THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

Vol. 59, Part 1 1 January 1947 No. 331

THE BREAK-UP OF LIQUID JETS

By A. C. MERRINGTON and E. G. RICHARDSON,
King's College, Newcastle-upon-Tyne

MS. received 10 *April* 1946

ABSTRACT. The mean drop size produced when a jet of liquid issuing from a nozzle nto the atmosphere, breaks up, has been investigated and experiments have been carried but both with fixed nozzles and with moving nozzles (attached to an aircraft) discharging backwards. A relation expressing drop size in terms of viscosity and the relative velocity between the jet and the surrounding air, applicable to both cases, is given.

At low velocities this relation ceases to apply and drop size reaches a limiting value. Factors influencing this limiting figure are discussed and information has also been obtained on the stability of large drops.

Measurements made on the length of the jet confirm the work of Tyler and Richardson but also disclose a hitherto unsuspected anomaly in the case of very viscous liquids.

Similar results were obtained when a jet issued into another liquid with which it did not mix, but showed the effect of the relative viscosity of the liquids on the break-up.

PART I—LIQUID INTO AIR

§1. INTRODUCTION

When a jet of liquid issues from a nozzle into the atmosphere, the jet eventually breaks up into drops under the action of disturbances of its equilibrium figure. The jet is observed to break up at a point which can be determined accurately, as long as the pressure behind the nozzle remains constant. While some previous work, both theoretical and experimental, exists on the length of this continuous portion of the jet, little quantitative work has been done on the size and behaviour of the drops subsequent to the break-up. In particular, there is a lack of experimental data on the effect of viscosity on the break-up of jets.

In the main there are three types of break-up. In the first the interplay of inertia and surface tension results in the jet becoming varicose. Rayleigh (1879) examined the conditions under which axial-symmetrical oscillations set up near the nozzle might increase in amplitude. In an inviscid jet he showed that a disturbance having a wave-length 4·4 times the diameter of the jet should grow fastest and eventually break up the jet into drops. Tyler and Richardson (1925) have taken photographs of this type of break-up on capillary jets. Rayleigh subsequently (1892) modified his theory to take account of the viscosity of the liquid, which naturally reduces the rate of growth of the optimum disturbance, whose wave-length in relation to the diameter of the jet remains unchanged.

In the second type of disturbance, the jet becomes sinuous and the resistance of the air to the passage of the humps becomes of more importance than surface tension. This type of break-up has been dealt with mathematically by Weber (1931) but is only amenable to semi-empirical treatment.

Since the air resistance to this type of motion increases rapidly with speed break-up will occur at a faster rate as the speed of efflux is increased, whereas the varicose form of disturbance has a rate of growth independent of the speed of efflux. When a liquid jet enters the air, both types of disturbance are equally possible but at low speeds the varicose form will break up the jet first. Since the growth coefficient of the "optimum" wave-length is constant, the continuous length of the jet will be proportional to the speed, as Smith and Moss (1917) showed, but at a critical speed the sinuous disturbance will grow at the same rate as the varicose; thereafter sinuosities will break up the jet first and the length will decrease with increasing speed.

After examining a number of photographs of high-velocity jets, both Haenlein (1931) and Ohnesorge (1937) postulated a third type of break-up in which the jet is disrupted, termed "atomization" (*zerstäubung*). Ohnesorge claims, too, critical velocities delineating regions for the three types of break-up. Littaye (1939) who has examined this question further, finds that the middle region of wave formation is somewhat indefinite, slight alterations in the conditions of the experiment being sufficient to convert this intermediate type of break-up into one of the other two.

Hence it may be concluded that at high velocities in excess of a critical value, break-up is controlled by viscous and inertia forces. Any expression deduced for this break-up should involve the characteristic factors which enter into viscous flow, i.e., a velocity, a drop diameter and the kinematic viscosity of the liquid.

(Although this work does not deal with gaseous jets or jets of liquid into the selfsame liquid, it may be remarked here that the sinuous type is the only one applicable to jets in which the interfacial tension is negligible, as, for example, in the sensitive flame.)

§ 2. THE MEASUREMENT OF DROP SIZE

Although some work of a qualitative nature has been carried out on the drops from jets delivered at high speed from narrow nozzles, in connection with engine oil injectors (Kuehn (1924), Scheubel (1927) and Schweitzer (1937)), exact measurements with proper control and measurement of the factors involved are lacking. The first part of this programme of work envisaged measurements of mean drop size, varying, in turn, speed of jet relative to air, nozzle diameter and viscosity of liquid, as being the fundamental factors likely to affect break-up in the atomization region, into which the majority of these experiments fall.

It was not possible to measure drop size immediately after break-up. Instead, the drops were allowed to fall on sheets of thin blotting paper (1 ft. diam.) at some distance. The stains, made clear by admixture of a dye in the liquid, were measured and the relation between stain-diameter and drop size established by shooting single drops of the same liquid from a pipette directly on to the paper; save for very small drops, this relationship was linear. With most liquids the relation was independent of impact velocity within the range of the investigation. Water was an exception, as large drops tended to splash on landing at high speed. As at a given jet speed the drops cover a range of sizes, it is necessary to define a "mean drop diameter". This may be done in a number of ways; the drop which occurs most frequently could be taken, or that size which is responsible for the greatest fraction of the total amount of liquid discharged. Actually it is the latter interpretation which is to be given to "mean drop size" in the results as plotted on the figures of this paper.

§ 3. MEAN DROP SIZE FROM STATIONARY AND MOVING NOZZLES

Drop-size distribution curves were obtained with a jet fired by gas pressure from a stationary nozzle on a cylinder directed vertically downwards in an enclosed tower and also from a nozzle horizontally discharging backward from an aircraft. In both cases the height of fall to the paper was about 50 ft. The jet velocity was calculated from the recorded pressure in the cylinder. In the tower experiments, overlapping of stains was prevented when the drops were insufficiently dispersed by swinging the board holding the filter paper on a bifilar suspension across the line of fire during emission. In the aircraft experiments, advantage was taken of a cross-wind to scatter the drops and many papers were spread out along and to one side of the line of flight. As it was intended that from measurements of the stains, drop sizes at break-up should be derived, it was necessary to consider the evaporation which takes place while the drops are falling. This calculation has been done in another connection (Richardson, 1946) and proves that with most liquids only the smallest drops (less than 0·25 mm. in diameter at this height) require correction for this cause. Exceptions to this statement must be made for an extremely volatile liquid like carbon tetrachloride, the uncorrected mean drop being, from this cause, too low.

The stains produced by the drops were counted in 1 mm. groups (0 to 1, 1 to 2, etc.) over a strip 3 inches wide passing through the centre of the pattern on the filter paper and hence the total number and the total volume occupied by each group calculated. When small drops predominated, the lowest group was further subdivided. The total volume occupied by all drops up to a particular size was plotted against mean diameter of the group. The curves (specimens of which are shown on figures 1 and 2) show that the distribution is symmetrical, so that the mean drop size could be taken to be that corresponding to 50 per cent of the total volume discharged.

A list of the liquids used and their physical properties follows:

Liquid	Density ρ (gm./cc.)	Kinematic viscosity ν (cm^2/sec.)	Surface tension σ (dynes/cm.)
Zinc chloride	1·76	0·10	40
Soap solution	1·00	0·020	30
Titanium tetrachloride	1·76	0·006	32
Carbon tetrachloride	1·60	0·006	25
Methylene chloride	1·33	0·004	30
Methyl salicylate	1·33	0·03	35
Chlorosulphonic acid	1·77	0·03	38
Water	1·00	0·012	73
Methyl salicylate (thickened)	1·34	7·0	25
Glycerine	1·26	10·0	64
Glycerine + 20 % water	1·21	1·0	45

The experiments with stationary nozzles indicated that the mean drop diameter d depends only on the viscosity of the liquid and on jet velocity V. In fact, d was found to be inversely proportional to V (i.e. to the relative velocity between the jet and the surrounding air). Two conical nozzles 0·15 and 0·25 inch diameter were used in addition to a 5/16-inch diameter sharp-edged orifice, and these gave identical results independent of nozzle size or shape. Some later tests with a small nozzle 0·030 inch diameter confirmed this.

As an example, figures obtained with two liquids, soap solution and glycerine, are given below:

Liquid	Jet velocity V (m./sec.)	Mean drop size d (mm.)	Vd/ν
Soap solution	25	1·10	14,000
	36	0·55	10,000
	57	0·38	11,000
	68	0·30	10,000
Glycerine	88	0·88	77
	109	0·73	79

In most of the experiments with the moving nozzles, the speed of the jet relative to the nozzle was greater than the aircraft speed, but in one instance it was less. In every case the mean drop size was found to depend only on the *relative* speed, V, of the jet to the air and on the viscosity of the liquid. Here again, d was inversely proportional to V. The nozzles used varied in diameter from 0·4 to 0·7 inch, and also gave identical results independent of nozzle size.

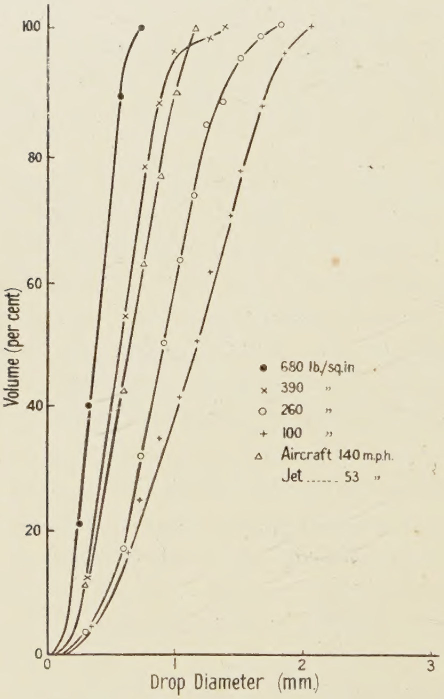


Figure 1. Zinc chloride.

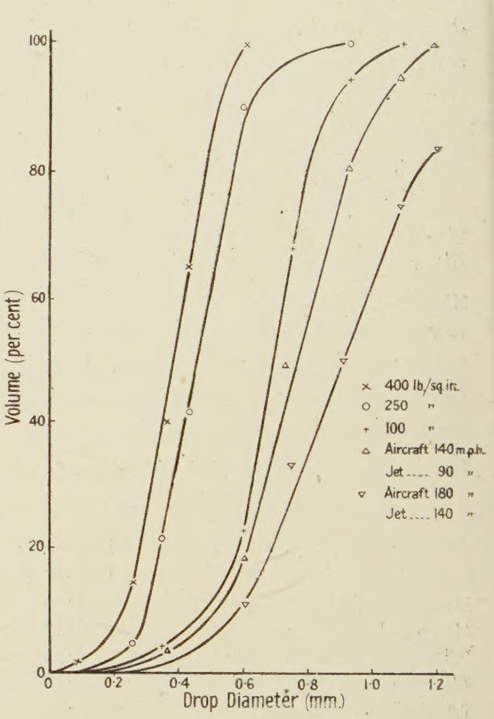


Figure 2. Titanium tetrachloride.

The surface tension of most of the liquids used is in the neighbourhood of 30 dynes per cm.; only water and glycerine have higher surface tensions, and these produced mean drop sizes expected from their respective viscosities and showed no effects which could be ascribed to surface tension. This result implies that at these speeds break-up is due to disruption of the jet.

On figure 3, Vd/ν is plotted against ν in log : log form. The results, for *both* stationary and moving nozzles, satisfy the empirical formula $Vd/\nu^{1/5} = 500$,* where V is the relative velocity between the jet and the surrounding air, d is the mean drop size, and ν is the kinematic viscosity of the liquid. Since static and moving nozzles give identical expressions for drop size, it is concluded that break-up in the "atomization" region is initiated by turbulence within the jet but is finally controlled by air friction on the jet surface. This is in agreement with the conclusion arrived at by Schweitzer (1937), who found that atomization could not be obtained without air friction, but to make air friction effective turbulence is essential.

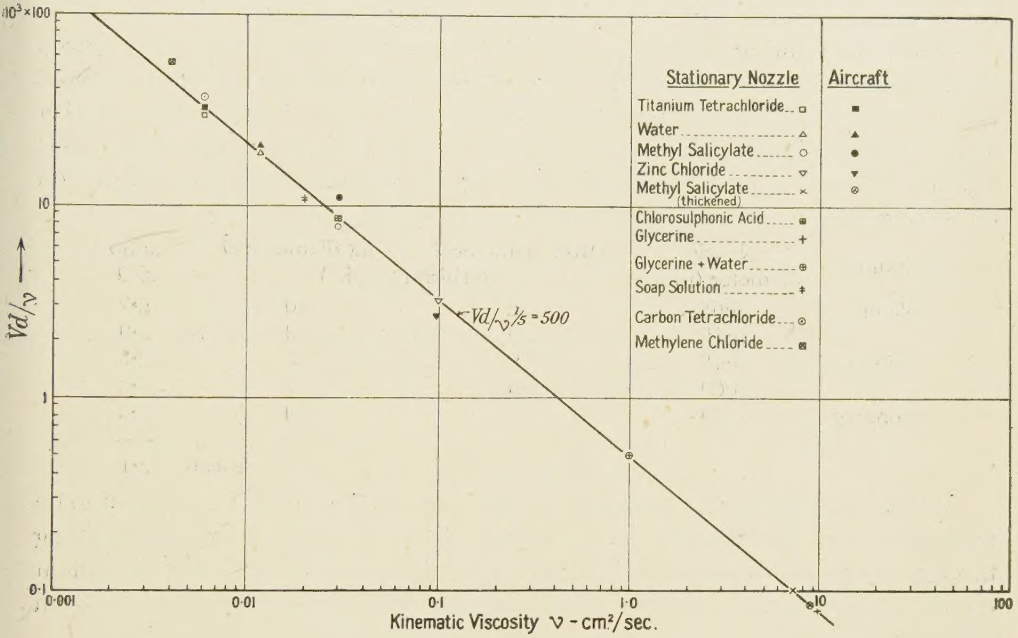


Figure 3.

§4. LIMITING DROP SIZE

It was postulated at the commencement of this paper that disruption of the jet is controlled by viscous and inertia forces and as such would predominate at high speeds. At lower speeds the varicose disturbance would be the dominant cause of break-up, while an intermediate type of break-up, in which the jet followed a sinuous motion, also occurs. We should, therefore, expect the relationship deduced from the above-mentioned experiments to cease to apply at low velocities. The varicosities then break the jet up into ovoid lumps which eventually resolve into drops comparable with the jet diameter and therefore with the nozzle diameter. Experiments (figures 4 and 5) show in fact that at low speeds the mean drop size reaches a constant value. The velocity at which this occurs can be read off the figures and compared with a value derived by Ohnesorge from an examination of photographs of jets. Ohnesorge's expression, which corresponds to the change-over from sinuous to the disruptive form of break-up, is

---

* This formula is, of course, dimensionally unsound, but could be corrected by including a term involving the kinematic viscosity of the surrounding medium. Since the latter was not varied in these experiments this term has been omitted.

$$\frac{\eta}{\sqrt{\rho\sigma D}} = 2{,}000\left(\frac{\eta}{V_0\rho D}\right)^{4/3},$$

where V_0 is the critical velocity, D the diameter of the nozzle, and η the viscosity of the liquid. Very viscous liquids are, however, broken up by the disruptive motion to a much lower speed than "thin" ones. For these the Ohnesorge criterion indicates too high a velocity for the change-over (figure 6).

Below the limiting velocity, the drops from narrow jets become, for a given liquid, uniform in size, so that the scatter of drop size apparent on figures 1 and 2 disappears. With fine nozzles (1 mm. diameter) the constant drop size reached at low velocities is roughly twice the nozzle diameter, with all but the very viscous liquids. This is in agreement with some figures obtained from an examination of photographs of varicose jets included in a paper by Tyler (1933).

Liquid	Nozzle diameter (mm.)	Drop diameter d	Jet diameter D	Ratio d/D
		(arbitrary units)		
Water	0·65	22	10	2·2
	0·45	20	10	2·0
Aniline	0·45	19	8	2·3
	0·60	16	7	2·3
Mercury	0·30	36	20	1·8
			Mean	2·1

When jets issue from *large* nozzles at low speeds, the drops are not found to be uniformly twice the diameter of the nozzle. This is because large drops falling through the air are unstable, dividing up into smaller units to an extent dependent on the path of fall and therefore on the distance of the collecting apparatus from the nozzle.

A drop of liquid in motion through another fluid differs in its behaviour from a solid sphere in that it may (*a*) be deformed, (*b*) have a circulation set up within itself by the shearing effect of the relative motion of the two fluids. These effects upset the stability of the drop, causing it to oscillate about the spherical shape and eventually to burst into fragments or, at least, into smaller drops.

Before bursting, the effects of deformation and circulation become apparent as an increase in the resistance, compared to that of a rigid sphere of the same diameter. Bond and Newton (1927) investigated this effect and showed that the parameter relevant to these changes was $(\rho - \rho')d^2g/\sigma$ where ρ and ρ' are the densities of drop (or bubble) and medium respectively. The value of this "Bond number" (B) was of the order unity when the rate of fall began to change. This value of B applied to air bubbles rising slowly in liquids; for liquid drops in a liquid medium it was higher—up to five in some cases. A few experiments (unpublished) have been done by Davies on the size of drops which will break up in the last 10 feet of their fall when released from heights of 30 or 40 feet. It appears that deformation ensues when B exceeds 0·4 and circulation within the drop becomes effective when $B = 1·5$. As these results are rather meagre and refer to small heights, it was thought desirable to make use of a tower of greater height (125 ft.) in the interior of which the drops were let loose. The drops were formed by the same method of aspirating some of the liquid up a graduated glass tube of between 1 and 4 mm. diameter and launching it by suddenly reversing the pressure applied to it. In this way drops up to 1 cm. diameter can be produced entire, although since they are not initially spherical, a certain amount of oscillation about the equilibrium form occurs before they settle down. In falling they are watched through a telescope at the top of the tower and the height at which they break up, if that occurs, calculated roughly by timing from release.

As the drops fell, they acquired an oscillation of figure which grew in amplitude until, if not stopped by the filter paper at the bottom of the tower, they broke either into two or three pieces, or else into a remaining drop of moderate size surrounded by a chaplet of tiny drops. The diameter of each drop was measured by its initial volume and checked by the diameter of the stain it produced. It was not possible to reproduce conditions exactly. Drops of the same size and liquid did not always break up at the same place in their fall, nor, if near the critical size, did both necessarily burst before reaching the bottom.

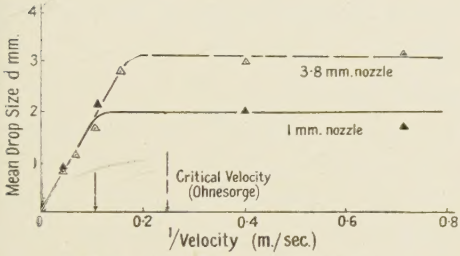


Figure 4. Carbon tetrachloride.

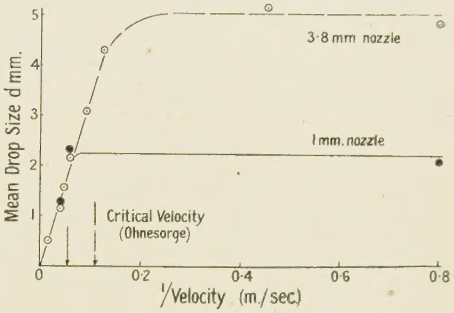


Figure 5. Water.

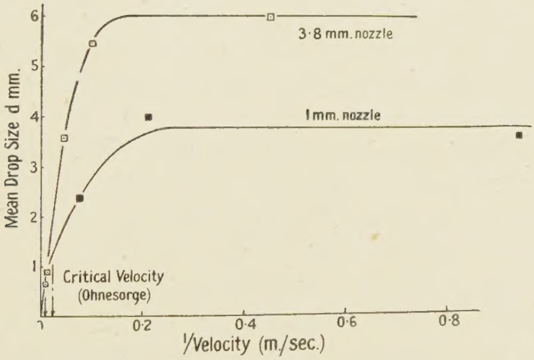


Figure 6. Glycerine.

By plotting d^2 against $100\sigma/\rho g$, that value of d (in mm.) was noted which best represented the critical size; hence the critical Bond number was deduced:

		d^2	d	B	d^2	d	B
Liquid	$100\sigma/\rho g$		from 125 ft.			from 50 ft.	
Dirty water	5·0	50	7·0	10	95	10	19
Distilled water	7·3	100	10·0	14	100	10	14
Carbon tretachloride	1·6	13	3·6	8	28	5	18
Tetrabromethane	1·55	16	4·0	10	25	5	16
Methyl salicylate	2·7	42	6·5	16	50	7	19
Do. (thickened)	2·9	72	8·5	25	80	9	28
Glycerine + 20% water	5·5	78	9·0	14	100	10	18

The mean values of B for "thin" liquids are 17 and 12 for 50 and 125 ft. respectively. The fact that a larger drop remains entire at the smaller height can be ascribed to the shorter time over which the disrupting forces can operate. The viscous liquid—thickened methyl salicylate—shows exceptionally high values of the critical B. This indicates a superior resistance to the tendency of shearing forces to produce circulation and is paralleled by Bond and Newton's observations of viscous oil drops in water.

To summarize this part of the work: at low velocities a change in the type of break-up may limit the mean size of drop attainable from a jet, and in such a case the mean size will no longer be independent of the diameter of the nozzle. In addition, the maximum drop size will in any type of break-up be limited by the size of the largest drop which can remain unbroken under the circumstances, even at high velocities and particularly with large nozzles.

§ 5. CONTINUOUS LENGTH OF JET

The shape of the curve relating length L of the continuous portion of the jet to velocity of efflux was first described by Smith and Moss (1917). There are two significant parts of the curve, (1) a straight portion: L/D proportional to $V\sqrt{\rho D/\sigma}$, (2) a hyperbolic portion. Tyler and Richardson in further experiments (1925) showed that the latter was governed by a relation $VL/\nu = \text{constant}$. This work was done with capillary nozzles ($<0{\cdot}6$ mm. diameter).

In order to extend the data on jet length provided by the 1917 and 1925 papers, measurements of jet length as well as of mean drop-size were made with the wide nozzles used in the present work.

Both Rayleigh (1892) and Weber (1931) have obtained expressions for the optimum rate of growth of "varicosities" on a cylindrical jet of *viscous* liquid. Viscosity, of course, superposes a damping term on the growth factor. Thus Weber's expression for the damping coefficient is

$$2\mu = -\frac{3\nu}{a^2}\epsilon^2 \pm \sqrt{\left\{\frac{9\nu^2}{a^4}\epsilon^4 + \frac{2\sigma}{a^3}(\epsilon^2-\epsilon^4)\right\}},$$

where a is the jet radius and $2\pi a/\epsilon = \lambda$, the wave-length of the disturbance. If ϵ is the same for optimum rate of growth, whether the jet be viscous or not (and Rayleigh considers this to be so) $\epsilon = 1/\sqrt{2}$ and

$$2\mu_{\text{opt.}} = \frac{1}{a^2}\left(-\frac{3\nu}{2} \pm \sqrt{\frac{9}{4}\nu^2 + \frac{\sigma a}{2}}\right).$$

Taking both negative signs we obtain the following values for -2μ with the radii 0·19 cm. and 0·05 cm. of the principal nozzles used in this research:—

ν cm²/sec.	-2μ	
	a=0·19 cm.	a=0·05 cm.
0·01	65	600
0·1	70	700
1·0	140	1,930
10·00	412	6,000

Now between viscosities 0·01 and 0·1 there is a negligible change in μ and therefore in L, but for $\nu = 1$ and $\nu = 10$ the lengths are respectively twice and seven times those with negligible viscosity.

Figure 7 shows the measurements of length with a nozzle of 3·8 mm. diameter. The straight portions of the characteristic curves for varicose jets of small viscosity can be made to coincide if L/D is plotted against $V\sqrt{\rho D/\sigma}$. Since all these results are for one nozzle and there is little difference in the parameter $\sqrt{D\rho/\sigma}$ it is apparent from an inspection of the figure that this no longer holds for viscous liquids. That this effect is not peculiar to the one nozzle was shown by similar measurements on 0·5, 1 and 2 mm. nozzles (figure 8), the smallest of which overlaps

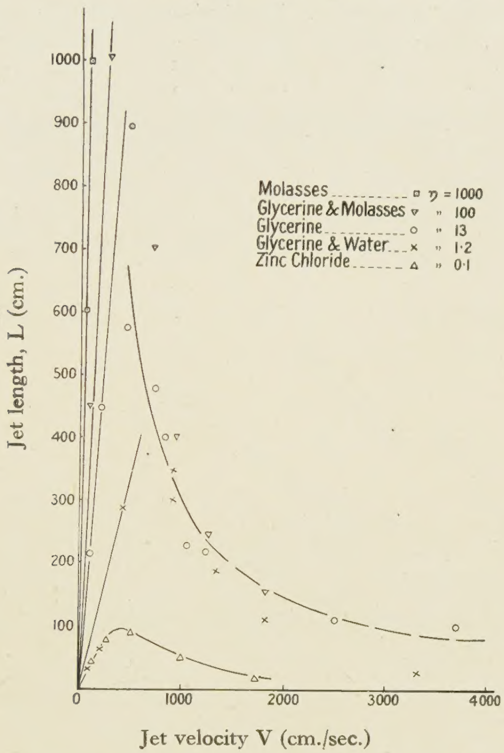


Figure 7. Results with jet 3·8 mm.

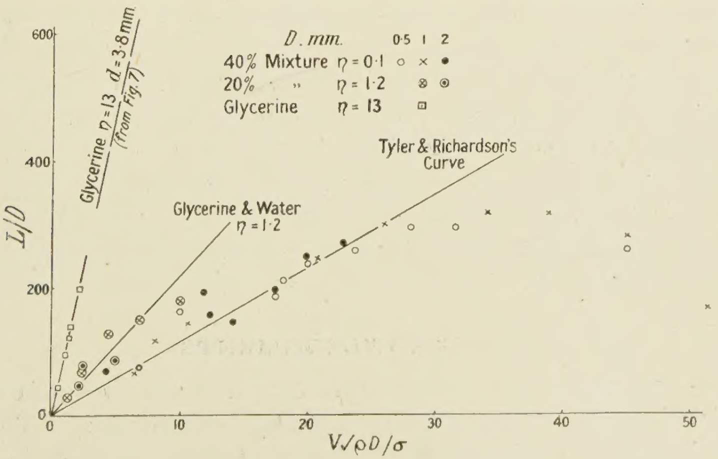


Figure 8. Capillary jets.

the range of nozzle size used by Tyler and Richardson, from whose figure 9 the dotted line on figure 8 is copied. Furthermore, the straight portion of the curves in figures 7 and 8 for $\nu=1$ and $\nu=10$ indicates lengths which are respectively about twice and seven times those with negligible viscosity. These figures agree well with those given above derived from Weber's theoretical expression for the damping coefficient.

Some experiments were also carried out on mercury and liquids exhibiting anomalous viscosity, one set of which, $1\frac{1}{2}\%$ solution of rubber in petrol, is shown with the mercury results on figure 9. Pressures up to 1000 lb. per sq. in. were used on mercury. These measurements are interesting because they confirm the observation of Tyler and Richardson that on contrast to all other liquids, jet length in mercury depended to a marked degree on the shape and method of manufacture of the nozzle.

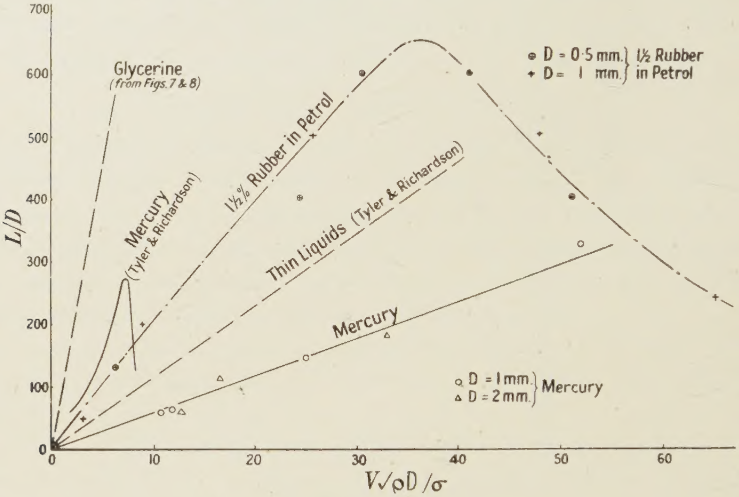


Figure 9.

The rubber solution shows a characteristic property of such fluids. Its position on figure 9 is appropriate to a viscosity of 2 or 3 poise, whereas its apparent value in an Ostwald viscometer (bore 0·5 mm.) was $\frac{1}{2}$ poise. Evidently the shearing forces in the motion of the liquid through the air are comparatively light, so that its effective viscosity is large.

Some photographs were taken of the mercury jets. One of these is interesting as it shows at various sections both varicose (figure 11) and sinuous motion (figure 10) at the same time. The varicose motion eventually persists and breaks up the jet (figure 12).

ACKNOWLEDGMENTS

This work was carried out by the Armament Research Department in collaboration with the Royal Aircraft Establishment, Farnborough. The authors are indebted to the Chief Scientific Officer, Ministry of Supply, and the Director, Royal Aircraft Establishment, for permission to publish this paper.

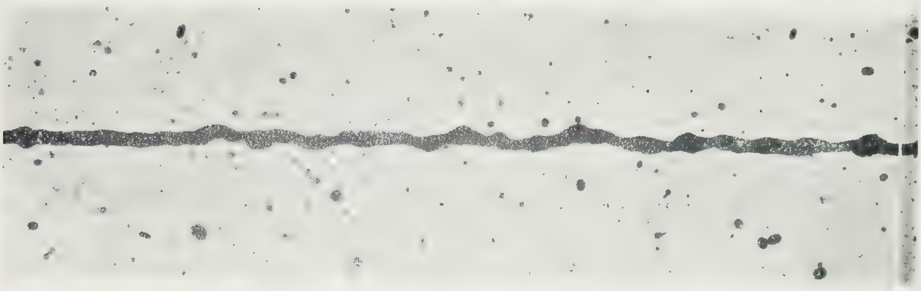


Figure 10. Sinuous motion in a mercury jet.
V=20·6 m./sec., nozzle diameter=2 mm.

Figure 11. Varicose motion.

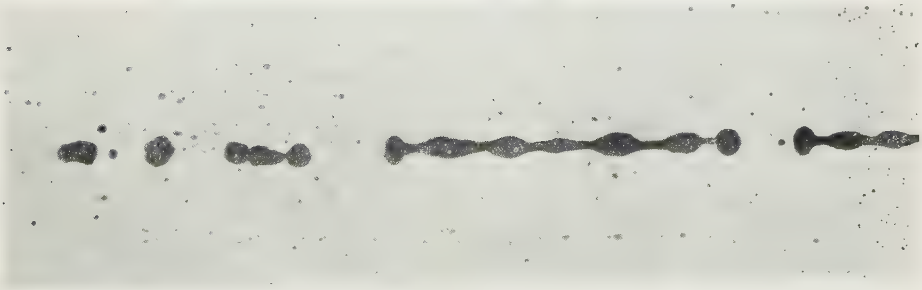


Figure 12. Varicose motion breaking up the jet.

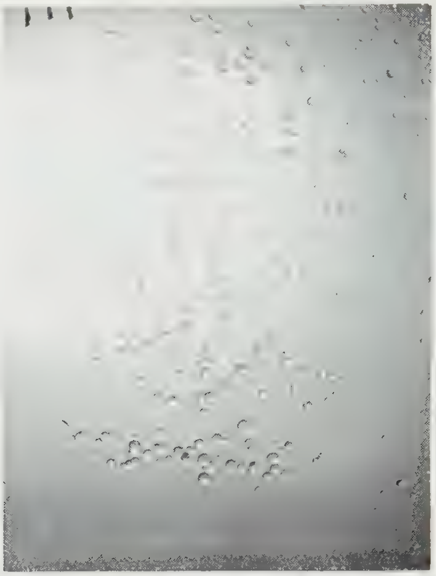


Figure 14 *b*.

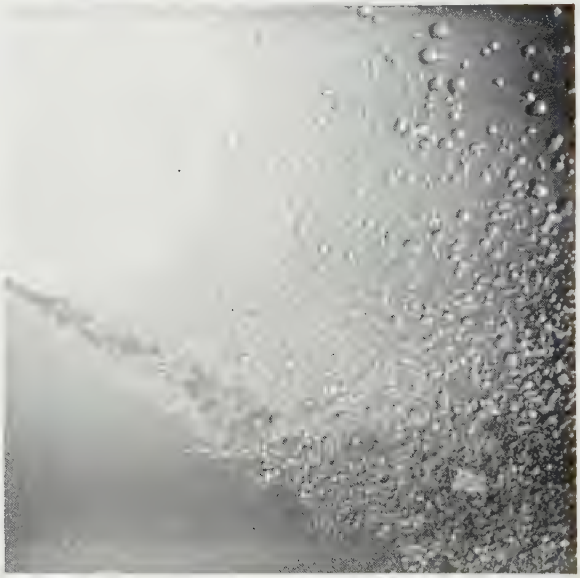


Figure 14 *a*.

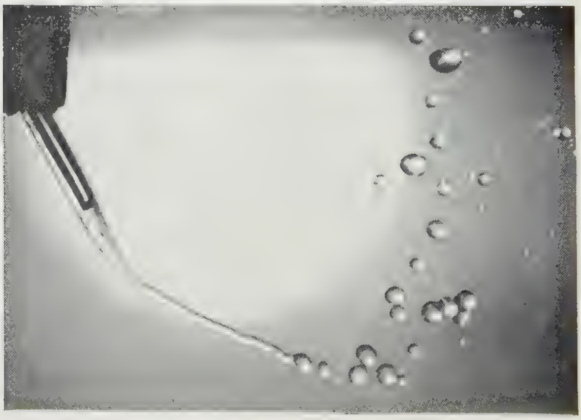


Figure 13.

PART II—LIQUID INTO LIQUID (IMMISCIBLE) JETS

§1. INTRODUCTION

In spite of its importance as a method of manufacturing emulsions in what s known as an "atomizer", we have not been able to find any quantitative data n the break-up of liquid jets into drops by injection into water or other liquids vith which they do not mix. Smith and Moss (1917) and Tyler and Watkin 1932) have extended their work on the lengths of the continuous portion of jets. oted in Part I, to cases in which one liquid is squirted from a capillary nozzle nto another liquid. The same general features were observed as in the case here the liquid emerges into air, i.e., a region in which change of form of the jet vithout change of direction caused it to break up into drops and another at higher fflux speeds when the jet broke up at the point where the amplitude of its sinuous notion became too great for its coherence to be preserved. If the jet passed into a ery viscous liquid, the growth of disturbances was hindered and the jet was nger than for the same liquid falling into air or water.

§2. EXPERIMENTAL WORK

The liquids forming the jets in the present research were aniline ($\nu = 0{\cdot}04$), enzene ($\nu = 0{\cdot}008$), paraffin ($\nu = 0{\cdot}023$) and various samples of fuel and lubricating il ($\nu = 0{\cdot}12$ to $0{\cdot}68$) squirted into water, and in the case of aniline, also into salt lution, contained in a glass-sided tank 4 ft. high by 18 inches square. One per ent of an emulsifier in the water prevented early recombination of the drops. he pressure used to eject them was read on a mercury manometer and afterwards rrelated with the velocity of efflux. An instantaneous photograph was taken the field of view some distance from the nozzle. At low speeds a fast shutter as used; at high speeds a discharge (arditron) lamp. Care had to be taken in noosing a limited portion of the field for analysis that drops were not dispersed tside the depth of focus (as this caused fogging) and were not in too great a ncentration (as this caused overlapping of images).

The photographs after development were projected, together with a superosed graticule, on a screen and the different sizes sorted to get summation curves milar to those of Part I. The limit of resolution was about ½ mm. (actual size).

Such curves were obtained for aniline, paraffin and oil, and benzene into water $\sigma/\rho = 6$, 40 and 35 c.g.s. respectively) and for aniline into brine ($\sigma/\rho = 12$) using ree circular nozzles of diameters 1·0, 1·5 and 2·75 mm. at speeds up to 8 m./sec. he critical velocities (cm./sec.) for the change of régime from varicose to sinuous, cording to Tyler and to Ohnesorge, are as follows:—

Liquids	Tyler's critical *V*			Ohnesorge's critical *V*		
	$D=0{\cdot}1$	$D=0{\cdot}15$	$D=0{\cdot}275$	$D=0{\cdot}1$	$D=0{\cdot}15$	$D=0{\cdot}275$
Aniline–water	46	38	28	70	50	19
Aniline–brine	65	53	40	98	70	26
Benzene–water	85	70	50	100	70	25
Paraffin–water	115	90	70	150	110	90

Comparison has been made between the Reynolds number for the mean dro diameter as observed and as calculated from figure 3 both directly and on th following basis: the abscissae of figure 3 were recalculated in terms of ν/ν where ν_0 is the dynamical viscosity of the surrounding medium (i.e. 0·15 for ai to which the figure refers) and the ordinate of Vd/ν read off, corresponding to th appropriate value of ν/ν_0 for the two liquids involved. These are given in th following table:—

Liquids	ν/ν_0	Reynolds number of mean drop		
		Obs.	Calc. for ν alone	Calc. for ν/ν_0
Aniline–water	3·0	1500	7000	700
Aniline–brine	3·5	1250	5000	500
Benzene–water	0·66	5000	27000	3000
Paraffin–water	2·0	3000	10000	1200
Oil–water	10·0	550	2000	250
,,	12·5	675	1900	240
,,	25·0	450	1500	200
,,	55·0	200	600	100

The observed values of Reynolds number for the mean drop are about twic those derived on a relative-viscosity basis but much below those appropriate t the absolute viscosity of the liquid forming the jet. Evidently, when the critica velocity is exceeded, friction with the surrounding medium has a greater influen on the break-up of the jet than friction within its own substance.

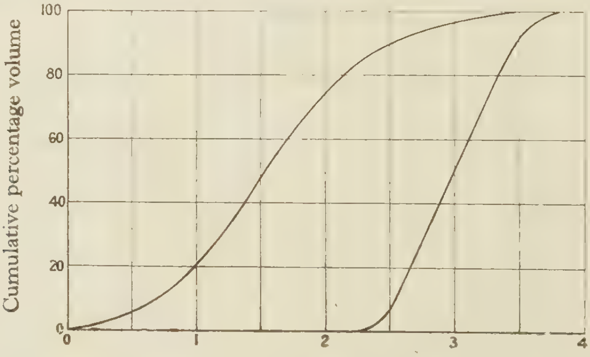


Figure 15. Drop diameter (mm.).

Below this, in the varicose régime, the drop size is governed by two additional factors, surface tension and nozzle size. The drops also tend to be more uniform in size. The difference can be seen on the specimen photographs reproduced. Figure 13 illustrates break-up in the neighbourhood of the critical velocity; figure 14 *a* shows the spread of the jet in the supercritical region, while figure 14 *b* illustrates a portion of the same jet sufficiently dispersed for analysis. The uniformity of size of drops at very low velocities is well seen in Tyler and Watkin's photographs of aniline–water jets, where the size is almost entirely governed by surface tension. When the speed of efflux is a negligible quantity, we have the well-known method of the "drop-weight" for the measurement of interfacial tension.

As a further comparison, two typical summation curves for drop size in the atomization (left-hand) and varicose (right-hand) régimes are shown on figure 15.

The effect of nozzle size is to limit the mean drop size at low velocities of efflux, so that if d were plotted against the reciprocal of V, we should get curves like those of figures 4 and 5. The effect on the liquid–liquid jets is shown in the following able :—

Benzene–water

	D=0·175			D=0·15			D=0·10			
V	95	160	220	100	140	170	87	123	215	→0
d	0·50	0·25	0·25	0·30	0·35	0·25	0·25	0·20	0·20	0·5
Vd	47	40	55	30	50	42	22	25	43	—

Aniline–brine

	D=0·275		D=0·15		D=0·10			
V	95	175	140	185	80	125	175	→0
d	0·37	0·25	0·18	0·25	0·30	0·35	0·28	0 37
Vd	35	45	24	46	24	45	50	—

In no case was break-up produced by disintegration of a drop once formed, s in certain cases of jets in air, the nozzles used being too fine to produce drops which were themselves unstable at the speeds at which they moved through the tationary liquid, so that the critical Bond number was never exceeded.

ACKNOWLEDGMENT

The work of Part II was carried out at King's College, Newcastle-upon-Tyne, y one of us (E.G.R.), who gladly acknowledges the help received from Mr. W. Bainbridge, research assistant.

REFERENCES

Bond and Newton, 1928. *Phil. Mag.*, **5**, 794.
Haenlein, 1932. *Forsch. Ing. Wes.*, **4**, 139.
Kuehn, 1924. *Der Motorwagen* (Berlin).
Littaye, 1939. *C.R. Acad. Sci., Paris*, **208**, 788 and 1705.
Ohnesorge, 1936. *Z. angew math. Mech.*, **16**, 355.
Rayleigh, 1879. *Proc. Lond. Math. Soc.*, **10**, 4.
Rayleigh, 1892. *Phil. Mag.*, **34**, 145.
Richardson, 1946. *Proc. Univ. Durham Phil. Soc.*, **10**, 394.
Scheubel, 1927. *Jahr. Wiss. Ges. Luftfahrt.*, 140.
Schweitzer, 1937. *J. Appl. Phys.*, **8**, 513.
Smith and Moss, 1917. *Proc. Roy. Soc.*, A, **93**, 373.
Tyler, 1933. *Phil. Mag.*, **16**, 504.
Tyler and Richardson, 1925. *Proc. Phys. Soc.*, **37**, 297.
Tyler and Watkin, 1932. *Phil. Mag.*, **14**, 849.
Weber, 1931. *Z. angew. math. Mech.*, **2**, 136.

THE LINES OF FORCE THROUGH NEUTRAL POINTS IN A MAGNETIC FIELD

By DAVID OWEN,
King's College, University of London

MS. received 3 *April* 1946

ABSTRACT. It is shown that, having precisely located a neutral point in a magnetic fi as the point of intersection of two directional loci, it is practicable to draw the lines of fo passing through the point. The theory of the field near a neutral point is treated fron geometrical point of view. Examples are shown of fields plotted when a bar magnet set parallel or at an angle to the earth's horizontal field.

---

§1. INTRODUCTORY

In a recent paper * a method of accurately locating neutral points in a magne field is described, any such point being determined as the point of intersecti of two directional loci which can be plotted with a small compass need The properties of the field near a neutral point thus assume a new practi interest. In particular, the lines of force passing through a neutral point can plotted, starting at the known point. If a small circle be drawn with the point centre it will be found that there are two diameters at right angles along which t compass needle will point, with one of its poles over or near the centre of the circ showing that there are two lines of force crossing orthogonally and reversi direction at the point.

It will be shown that in general the lines of force through a neutral point cons of an isolated line in a certain direction and an infinite number of lines in the pla at right angles. If the direction of the field along the first line is towards t neutral point the directions of all the lines in the orthogonal plane are away fro the point. The tangent to the first line, at the neutral point, may appropriate be termed the principal axis of the field near the neutral point. Usually t selected plane in which a field is plotted contains that axis, and in such case accordingly, two lines of force only will be found passing through the point in t plane of plotting.

A theoretical treatment of the field in the vicinity of a neutral point is develope from a geometrical point of view, attention being directed to lines of force an field intensities as well as to equipotential lines and surfaces.

Two examples in illustration are given of the fields plotted when a bar magn is set horizontally in the earth's field, first with its axis parallel, then at an angle, that field. Some interesting properties of the field due to the superposition the field of a circular current on the earth's field are also pointed out.

---

* Owen, *Proc. Phys. Soc.*, **57**, 294, 1945.

§2. THEORY

The mathematical investigation of neutral points and lines of neutral points has usually been given with exclusive reference to electric fields.* Neutral lines, or lines of equilibrium, are of common occurrence on the surfaces of charged conductors, and whole regions of space may occur in which the potential is constant. Such phenomena have no counterpart in magnetic fields since, unlike lines of electric induction, lines of magnetic induction are always closed curves. Thus the points of emphasis in the two kinds of field are different. The fact that magnetic lines of force can be readily plotted lends a practical interest to the study of the magnetic field.

In place of the usual analytical treatment a geometrical approach is here proposed, attention being directed to lines of force and field intensities as well as to the equipotential lines and surfaces which appear to be inevitable at the outset. Take any plane containing a neutral point, and choose rectangular axes of x and y in the plane, with origin at the point. Measuring potentials from a value taken as zero at the neutral point, the magnetic potential V at any point (x, y) can be represented by the length of an ordinate erected perpendicular to the plane. A continuous "surface of potential" (not to be confused with an equipotential surface) can thus be generated which touches the plane at the origin. The curves of section of this surface by planes parallel to the x, y plane, when projected on that plane, are equipotential lines. Close to the origin all such surfaces become approximately of the second degree and, as is shown in treatises on solid geometry, belong to one of three classes: the elliptic paraboloid, the hyperbolic paraboloid, and the parabolic cylinder.

The first of these, the elliptic paraboloid, lies wholly an one side of the x, y plane, and the corresponding equipotential lines are similar ellipses.

The second, the hyperbolic paraboloid, is a saddle-back or anticlastic surface, ying partly above and partly below the x, y plane. The curves of section by planes above the reference plane are similar hyperbolas and the equipotential ines obtained by projection are of one sign of V; while the sections below the plane yield on projection the conjugate hyperbolas, with the opposite sign of V. The reference plane itself cuts the surface in a pair of straight lines through the origin, namely, the common asymptotes $V=0$ of both sets of hyperbolas.

The third class of surface, the parabolic cylinder, lies wholly on one side of the reference plane, and the sections parallel to the plane are pairs of straight lines. Such a surface would lead to a field in which all the equipotential lines were parallel, and therefore cannot represent the essentially non-uniform field near a neutral point. This consideration eliminates the parabolic cylinder as a possible form of surface of potential.

The form of the equipotential surfaces near a neutral point can now be determined. Sections by any plane through the point must, as we have shown, be either ellipses or hyperbolas. The ellipsoidal form of surface is excluded by the consideration that it would imply the existence of a field with lines of force all converging to or diverging from the neutral point, in contradiction to Gauss's heorem as applied to magnetism. The hyperbolic form of equipotential surface remains as the unique general solution. The equipotential surfaces near neutral point accordingly consist of a system of similar hyperboloids of two sheets with potentials of one sign, and of the conjugate hyperboloids of one sheet, with potentials of the opposite sign, the two sets being separated by the asymptoti double cone $V=0$. The form of the members of such a system is indicated i figure 1.

---

* See Maxwell's *Electricity and Magnetism*, Vol. I, Chaps. VI and VII, and Jeans's *Electricity and Magnetism*, Chap. II.

The principal axes of the system may now be adopted as the rectangular axes of x, y and z, with axial lengths a, b and c respectively. The general equatio of the whole system is

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}-\frac{z^2}{c^2}=\pm 1, \qquad \ldots\ldots(1)$$

the positive sign on the right-hand side applying to the 2-sheet hyperboloids, th negative sign to the 1-sheet hyperboloids. As a, b and c change values from zer upwards, their ratios remaining constant, the successive equipotential surfaces are described.

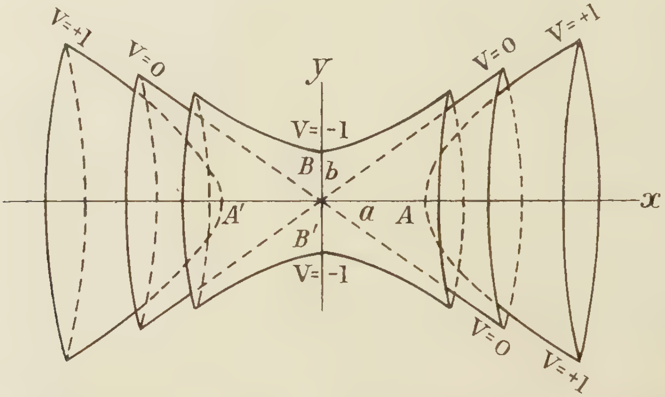


Figure 1. Equipotential surfaces close to a magnetic neutral point, showing one 2-sheet hyperboloid ($V=1$) and one 1-sheet hyperboloid ($V=-1$), separated by the asymptotic con ($V=0$). The axes of x and y (lengths $OA=a$ and $OB=b$) are taken in the plane of th paper. The axis of z (length c) is not shown.

It can at once be seen that the lines of force through a neutral point, being orthogonals to the equipotential surfaces, consists of one line along the x or a axis and an infinite number of lines in the perpendicular plane. The lines of force in that plane can be shown to be curves satisfying the equation $z=Ky^n$, where $n=b^2/c^2$ and K takes all values from zero to infinity.

The x-axis can thus be regarded as the principal magnetic axis of the field at the neutral point. Lines of force plotted in any plane containing that axis (as in the two practical examples given below) will show only two lines of force through the neutral point, crossing at right angles at the point; if the direction of the field is towards the neutral point along one line, it is away from it along the other.

Since the surface of potential with respect to any plane through a neutral poin is of the second degree and touches the plane at that point, the potentials at the extremities of the axes of any equipotential hyperboloid may be denoted by Aa^2, $-Bb^2$ and $-Cc^2$, where A, B and C are positive constants. Thus

$\partial^2 V/\partial x^2 = 2A$, $\partial^2 V/\partial y^2 = -2B$, and $\partial^2 V/\partial z^2 = -2C$. Applying Laplace's equation

$$\partial^2 V/\partial x^2 + \partial^2 V/\partial y^2 + \partial^2 V/\partial z^2 = 0$$

it follows that $A - B - C = 0$, thus imposing a relation between a, b and c, namely

$$1/a^2 - 1/b^2 - 1/c^2 = 0. \qquad \ldots\ldots(2)$$

In general, therefore, both b and c must be greater than a, so that the field gradient is steepest along the x-axis. Two special cases occur: first when $b = c = a\sqrt{2}$; next when $a = b$, $c = \infty$. In the first case the hyperboloids become surfaces of revolution about the x-axis. The two equipotential lines $V = 0$ in any plane through the x-axis make angles $\tan^{-1}\sqrt{2}$ with the axis; and the field-gradient along that axis is twice that along the y- and z-axes, and of the opposite sign. This case is illustrated in figure 2. In the second case the equipotential surfaces become hyperbolic cylinders, the equipotential lines $V = 0$ in the plane $z = 0$ make angles of 45° with the x- and y-axes, the field-gradients are equal but of opposite sign along those axes, while the z-axis is tangent to a circle of neutral points. This case is illustrated in figure 3.

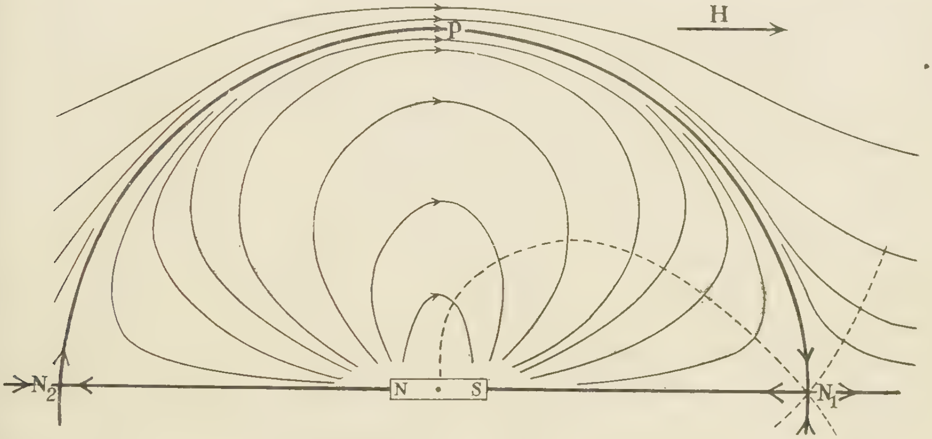


Figure 2. Field (one half, to half-scale) of a cobalt-steel magnet 1″ long, ¼″ in diameter, axis parallel to H, S-pole to the north. The lines of force through the neutral points are drawn of double thickness. The dotted lines are the plotted equipotentials through one of the neutral points, and make angles with the axis agreeing well with the theoretical value $\tan^{-1}\sqrt{2}$.

§ 3. EXPERIMENTAL

The two practical examples of plottings of fields containing neutral points relate to a simple bar magnet set with its axis horizontal in the earth's field. As the vertical component of that field has no influence on plotting in a horizontal plane, the lines of force obtained are of course identical with those due to the superposition of a uniform field H on the field of the magnet. The pair of lines of force crossing orthogonally at the neutral point are drawn of double thickness.

In the first example (figure 2) the magnet is set with its axis parallel to H, with S-pole to the north. The principal line of force, through both neutral points, is along the axis of the magnet. The second line N_1PN_2 also contains both neutral points; with the complementary half (not shown) it forms a nearly circular closed curve, a little flattened in the direction of the equatorial line of the magnet. It is interesting to note that were the magnet replaced by a doublet of equal magnetic moment this line would become a circle; and that at any point on the circle, taking the doublet as origin and measuring θ from the axis, the field intensity would be $\frac{3}{2} H \sin \theta$.

The second example illustrates a more general case, the bar magnet being set with axis perpendicular to H.

Interesting examples, easy to plot experimentally, also occur when a circular current is set with plane vertical and axis parallel to H. If the two fields oppose along the axis the pair of neutral points coincide at the centre of the coil for a certain value of the current. For current values above this the neutral points separate along the axis, while for values below it they separate along the diameter. On reversing the current the neutral points are always on a diameter and outside the coil.

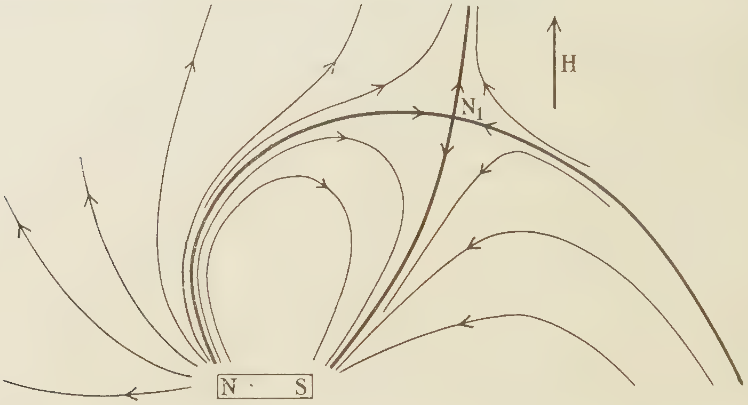


Figure 3. Upper half of field (to half-scale) due to the same cobalt-steel magnet placed with its axis perpendicular to H, N-pole to the west.

When the neutral points fall on the axis the same pair of lines of force pass through both neutral points, one coinciding with the axis, the other embracing the coil; and the pair of equipotential lines $V = 0$ make angles $\tan^{-1} \sqrt{2}$ with the axis.

When, however, the neutral points lie on a diameter a looped line of force passes through each neutral point, cutting itself orthogonally there. One of the equipotential lines through either neutral point is along a diameter, the second is parallel to the axis, so that both lines make angles of 45° with the lines of force through the point.

SENSITIVITY AND IMPEDANCE OF ELECTRO-ACOUSTIC TRANSDUCERS

By P. VIGOUREUX,

Torpedo Experimental Establishment, Greenock

MS. received 29 *May* 1946

ABSTRACT. The equivalent electric circuits of piezoelectric and magnetostriction electro-acoustic transducers are derived in terms of their mechanical constants, and the potential difference across the output terminals of a transducer placed in a given sound field is calculated in terms of the pressure of the undisturbed field and the radiation resistance of the transducer.

It is shown that, with certain reservations, the radiation resistance can itself be estimated from purely electrical measurements of the impedance of the transducer in air and in water over a range of frequencies near resonance, or even at a few selected frequencies.

The construction of impedance and admittance diagrams is explained, and it is shown how the frequency of maximum acoustic output of transducers can be obtained from electrical measurements without any acoustic measurements.

§ 1. INTRODUCTION

THE measurement of the intensity of supersonic fields in various media, e.g. in water, is usually effected by means of a calibrated hydrophone, i.e. an electro-acoustic transducer which gives a voltage proportional to the pressure r particle velocity of the field in which it is placed. It is not easy to make a hydrophone having the same sensitivity over a wide range of supersonic frequencies, but here are methods of obtaining the calibration over the range.

In many cases, however, the measurement of the supersonic field is required nly in connection with an installation comprising an electro-acoustic transducer f a different kind, used primarily, say, for under-water signalling or for submarine letection, and, with certain limitations, it is possible to make use of the same ransducer for measurement of the sound field, thus avoiding the necessity for an uxiliary hydrophone. This transducer operates only at or near its resonance requency and its vibrating face has dimensions large compared with the wave-ength; it is suitable only for measurement of supersonic fields of frequencies ear its operating frequency, but these are also in general the only fields of interest or the practical use of the transducer. The only data required for this measurenent are the face area and the radiation resistance of the transducer, and the only bservation required is that of the potential difference which appears at the erminals of the transducer as a result of the sound field in which it is placed.

The radiation resistance can itself be derived from purely electrical measurenents of the impedance of the transducer in air and water over a range of freuencies near resonance, or even at a few selected frequencies, provided it can be scertained that the mode of vibration is the same in air as in water.

The theory underlying those measurements is given below for piezoelectric nd magnetostrictive transducers.

§2. MECHANICAL IMPEDANCE OF TRANSDUCER

The only accessible parts of the transducer, as far as the measurement considered here are concerned, are its vibrating face and its two terminals A and B For more generality let an external impedance Z_e be connected across A, B.

Denote by $\mathscr{Z}_{\mathrm{T}}$ the mechanical impedance of the transducer under thes conditions, i.e. the quotient of the force F applied to the face of the transduce and the resulting velocity u of the face.

When the transducer is immersed in an elastic medium, displacement of th face is resisted not only by the impedance $\mathscr{Z}_{\mathrm{T}}$ of the transducer, but also by a additional mechanical impedance $\mathscr{Z}_{\mathrm{W}}$ which depends on the area A of the face an on the frequency; the velocity of the face is, therefore, inversely proportional t $\mathscr{Z}_{\mathrm{W}}+\mathscr{Z}_{\mathrm{T}}$. It must also be proportional to the free field pressure p, and is, therefore, equal to $mp/(\mathscr{Z}_{\mathrm{W}}+\mathscr{Z}_{\mathrm{T}})$, where m is independent of $\mathscr{Z}_{\mathrm{T}}$ but may depend o the face area and on frequency. But when $\mathscr{Z}_{\mathrm{T}}$ is equal to $\mathscr{Z}_{\mathrm{W}}$, the sound field i front of the face is not disturbed by the transducer,* for the impedance would no be altered by it; accordingly, in that case, the velocity of the face is the same a the free field particle velocity, viz. $p/\rho c$; hence

$$\frac{mp}{2\mathscr{Z}_{\mathrm{W}}}=\frac{p}{\rho c} \quad \text{or} \quad m=\frac{2\mathscr{Z}_{\mathrm{W}}}{\rho c}.$$

The velocity can therefore be written

$$u=\frac{2p\mathscr{Z}_{\mathrm{W}}}{\rho c(\mathscr{Z}_{\mathrm{W}}+\mathscr{Z}_{\mathrm{T}})}.$$

In the case when the dimensions of the transducer face are large compared wit the wave-length, the "plane wave damping" condition is approached, and th impedance $\mathscr{Z}_{\mathrm{W}}$ is nearly equal to the pure resistance ρcA. In this case the abov formula reduces to

$$u=\frac{2pA}{\rho cA+\mathscr{Z}_{\mathrm{T}}}. \qquad \text{......(1)}$$

Expressions must now be obtained for the impedance $\mathscr{Z}_{\mathrm{T}}$ for various types o transducers. These types can be divided into two classes, one of which include the electrostatic and piezoelectric types, the other the electromagnetic an magnetostriction types.

Let the electrical impedance from A to B, i.e. the quotient of voltage applie to AB, and current in AB when the face of the transducer is clamped, be Z, and le the mechanical impedance, i.e. the quotient of force applied to the face an velocity of the face when the terminals A, B are on open circuit, be $\mathscr{Z}$. In th absence of electrical input, if a force F applied to the face produces a velocity u the total power input is the product $F \cdot u$, and when electromechanical transfe of energy is allowed to occur, this power must be equal to the sum of the power dissipated mechanically and electrically.

In the piezoelectric case the quantity of electricity liberated at the electrodes i proportional to face displacement, in other words, face velocity u produces in th impedance Z_p formed by Z and Z_e in parallel a current ku, where k is a constant of the transducer independent of frequency. Since the electrical power is $Z_p i \,.\, i$, the power equation is

---

* Actually there is the additional requirement that the acoustic impedance per unit area b uniform over the face of the transducer, otherwise the field is distorted. This additional conditio is approached when the face dimensions are large compared with the wave-length. The absenc of a rim and other stationary supports near the face is also implied, as they would cause distortion

$$F \,.\, u = \mathscr{Z} u \,.\, u + Z_p ku \,.\, ku,$$

from which, after division by u^2, the total impedance is found to be

$$\mathscr{Z}_{\mathrm{T}} = \frac{F}{u} = \mathscr{Z} + k^2 Z_p. \qquad \ldots\ldots(2)$$

In the magnetostriction case, on the other hand, the magnetic flux linking the winding is proportional to face displacement; in other words, velocity u produces in the loop of total impedance Z_s formed by one or more parts of Z (e.g. inductance and series resistance to the winding), in series with Z_e and the remaining parts (e.g. shunt resistance of the winding), an e.m.f. hu where h is a constant of the transducer independent of frequency. Since the electrical power is $e \,.\, e/Z_s$, the power equation is

$$F \,.\, u = \mathscr{Z} u \,.\, u + hu \,.\, hu/Z_s,$$

from which the total impedance is found to be

$$\mathscr{Z}_{\mathrm{T}} = \frac{F}{u} = \mathscr{Z} + h^2/Z_s,. \qquad \ldots\ldots(3)$$

§3. SENSITIVITY OF PIEZOELECTRIC RECEIVER

In most cases the quantity observed or used is not the face velocity but the electric potential v to which it gives rise across AB. In the piezoelectric case this voltage is equal to the current through Z_p multipled by Z_p. The current is equal to ku, and by insertion of the value of u from (1), substitution of the value of $\mathscr{Z}_{\mathrm{T}}$ from (2) and abbreviation of the motional impedance $(\rho cA + \mathscr{Z})/k^2$ by J, it is found that

$$\begin{aligned} i = ku &= \frac{2pkA}{\rho cA + \mathscr{Z}_{\mathrm{T}}} \\ &= \frac{2pkA}{\rho cA + \mathscr{Z} + k^2 Z_p} \\ &= \frac{2pA}{k} \cdot \frac{1}{Z_p + J}, \qquad \ldots\ldots(4) \end{aligned}$$

from which it appears that the representation of the piezoelectric receiver can be completed by addition to Z_p of an electric curcuit J in which there is an e.m.f. e equal to $2pA/k$. If a, b, g are the mass, damping and restoring constants of the transducer, $\mathscr{Z}$ is equal to $aD + b + g/D$, where D is the time differential operator. If this value of $\mathscr{Z}$ be inserted in the formula for J, it is found that J is equal to $ND + S_i + 1/KD + \rho cA/k^2$, where N, S_i and K are written for a/k^2, b/k^2 and k^2/g respectively, and represent the equivalent series inductance, internal resistance and capacitance of the transducer. By analogy the term $\rho cA/k^2$ must be identified with the radiation resistance S of the transducer; this relation gives

$$k = \sqrt{\frac{\rho cA}{S}}, \qquad \ldots\ldots(5)$$

and, therefore,

$$e=\frac{2pA}{k}=2p\sqrt{\frac{AS}{\rho c}}, \qquad \ldots\ldots(6)$$

and the circuit of figure 1, into which is injected an e.m.f. e independent of frequency, gives an electrical representation of the behaviour of the receiver. In that circuit, the impedance Z_p comprises the capacitance c of the transducer and the external impedance Z_e in parallel.

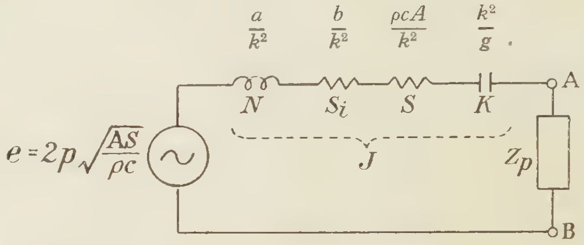


Figure 1. Electrical equivalent of piezoelectric receiver.

§4. SENSITIVITY OF MAGNETOSTRICTION RECEIVER

In the magnetostriction case, the voltage v across AB is equal to that produced by an e.m.f. hu in the winding L which forms part of the series circuit Z_s. This e.m.f. hu produces a current i in the winding and the rest of Z_s, and the terminal voltage is obtained by multiplication of i by the lumped impedance of all elements between A and B excluding that containing L. Insertion of the value of u from (1), substitution of the value of $\mathscr{Z}_{\mathrm{T}}$ from (3), and abbreviation of the motional admittance $(\rho cA+\mathscr{Z})/h^2$ by G yields

$$\begin{aligned} i=\frac{hu}{Z_s} &= \frac{2phA}{(\rho cA+\mathscr{Z}_{\mathrm{T}})Z_s} \\ &= \frac{2phA}{\left(\rho cA+\mathscr{Z}+\dfrac{h^2}{Z_s}\right)Z_s} \\ &= \frac{2pA}{h}\cdot\frac{1}{1+GZ_s}, \qquad \ldots\ldots(7) \end{aligned}$$

from which it appears that the representation of the magnetostriction receiver can be completed by addition across Z_s of an admittance G, the whole being fed with a current i_0 equal to $2pA/h$. Since as before the mechanical impedance $\mathscr{Z}$ may be represented by $aD+b+g/D$, where D is the time differential operator, the admittance G is equal to $KD+\frac{1}{S_i}+\frac{1}{ND}+\frac{\rho cA}{h^2}$, where K, S_i and N are equal to a/h^2, h^2/b and h^2/g respectively, and represent the equivalent parallel capacitance, resistance and inductance of the transducer. By analogy, the resistance $h^2/\rho cA$ must be identified with the radiation resistance S of the transducer. This relation gives

$$h=\sqrt{\rho cAS} \qquad \ldots\ldots(8)$$

and

$$i_0=\frac{2pA}{h}=2p\sqrt{\frac{A}{\rho cS}}, \qquad \ldots\ldots(9)$$

and the circuit of figure 2 fed with a current i_0 independent of frequency gives an electrical representation of the behaviour of the receiver.

In general the impedance Z consists of the inductance L of the winding and its resistive component. As the loss in the winding is chiefly due to eddy currents, a shunt resistance is a better representation of it than a series resistance, because the former gives a phase angle proportional to frequency as required by eddy-current loss. This shunt resistance r must be inserted across $L+G$ in the electrical representation. Consideration of the case when Z_e is infinite shows that it must not be inserted across L alone, because it would then have no influence on the terminal voltage, which in reality it tends to reduce, since it represents loss by eddy currents.

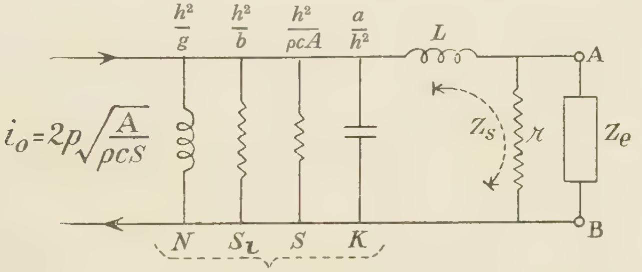
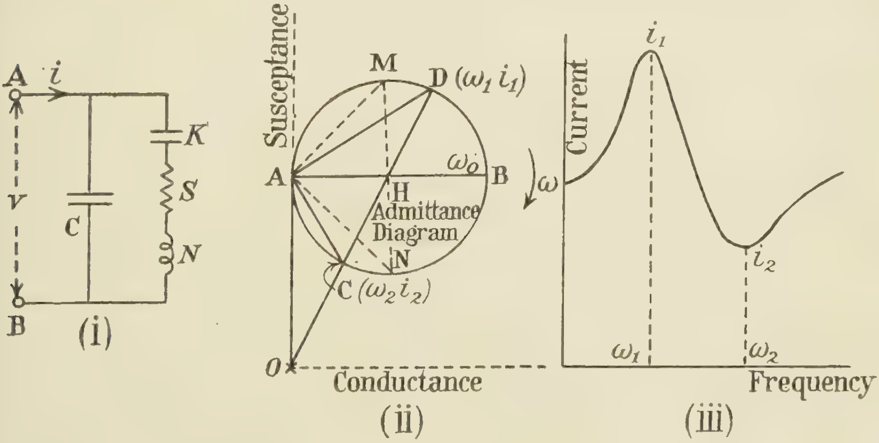


Figure 2. Electrical equivalent of magnetostriction receiver.

§ 5. IMPEDANCE OF PIEZOELECTRIC TRANSDUCER

With the reservations made in § 1, the radiation resistance S can be determined by purely electrical measurements.

In the case of the piezoelectric transducer, which is a fairly high-impedance device, there is no difficulty in maintaining a constant voltage across the terminals. The current observed is then proportional to admittance. In the absence of an external impedance Z_e, the impedance Z_p of figure 1 reduces to that of the condenser C, as in figure 3 (i), and by measuring admittance at frequencies well above



$$\omega_0=\sqrt{\omega_1\omega_2}\ \text{approx.},\quad S=\frac{v}{i_1-i_2},\quad C=\frac{\sqrt{i_1i_2}}{v\omega_0},$$

$$Q=\frac{1}{SK\omega_0}=\frac{\sqrt{\omega_1\omega_2}}{\omega_2-\omega_1}\cdot\frac{1}{2}\left(\sqrt{\frac{i_1}{i_2}}+\sqrt{\frac{i_2}{i_1}}\right)\ \text{approx.}$$

Figure 3. Analysis of piezoelectric transducer.

and well below resonance, the admittance of the condenser alone can be evaluated at any frequency by proportion, and subtracted vectorially from the admittance of the whole network. If the remaining admittance is plotted vectorially, a circle of diameter $1/(S+S_i)$ is obtained, and the Q value $1/(S+S_i)K\omega_0$ is equal to $\omega_0/\Delta\omega$ where $\Delta\omega$ is the frequency difference between the two quadrantal positions, i.e. the two positions M and N, figure 3 (ii), for which the vectors are at 45° to the diameter. The constants of the transducer are thus completely determined.

In order to separate S from S_i, it is necessary to perform measurements with the transducer in air and in water or other medium in which it is used. The value of ρc for air is so much smaller than it is for water that in many cases it can be neglected altogether, but anyhow a correction can be applied. If the linearity of the transducer has not previously been established, the measurement in air must be effected with a voltage which gives the same amplitude of vibration as in water; in other words, the difference between the maximum and minimum currents must be the same in each case. Even then, subsidiary vibrations, which occur more readily in air than in water, may vitiate the results.

When the Q value is high and speed of measurement is desired rather than great accuracy, it is sufficient to observe only the two maximum and minimum currents i_1, i_2, for an applied voltage v, and the corresponding frequencies ω_1, ω_2 (figure 3 (iii)). It is shown in the Appendix that the following approximations hold:

$$\omega_0 \simeq \sqrt{\omega_1\omega_2}, \qquad \ldots\ldots(10)$$

$$S = \frac{v}{i_1 - i_2}, \qquad \ldots\ldots(11)$$

where, for brevity, S is written for the sum of the internal and radiation resistances,

$$C = \frac{\sqrt{i_1 i_2}}{v\omega_0}, \qquad \ldots\ldots(12)$$

$$Q = \frac{1}{SK\omega_0} = \frac{\sqrt{\omega_1\omega_2}}{\omega_2 - \omega_1} \cdot \frac{1}{2}\left(\sqrt{\frac{i_1}{i_2}} + \sqrt{\frac{i_2}{i_1}}\right). \qquad \ldots\ldots(13)$$

The analysis of the piezoelectric transducer into its components is thereby complete. It must, however, be emphasized that this method, whilst valuable for its rapidity, is not comparable in accuracy with the more elaborate measurement of admittance at a number of frequencies and plot of the circle diagram, especially when the Q value is low.

§ 6. IMPEDANCE OF MAGNETOSTRICTION TRANSDUCER

As the magnetostriction transducer is a low-impedance device, a convenient way of measuring its electrical performance is to send a known current through it and to measure the voltage at the terminals for a number of frequencies. The impedance in series with L in figure 2 reduces in this case to the shunt resistance r of the winding L as in figure 4 (i). This measurement gives the impedance of the whole device. To obtain the motional impedance it is necessary to subtract the impedance of the winding. This impedance can be evaluated at any frequency by proportion if readings at two frequencies, one well below, the other well above resonance, are taken. The ends of the motional impedance vectors should lie on a circle of diameter approximately equal to S, where for brevity S is taken as the resultant of the internal and radiation resistances of figure 2. The Q value, which in this case is the resistance multiplied by $K\omega_0$, is equal to $\omega_0/\Delta\omega$, where $\Delta\omega$ is the difference between the quadrantal frequencies. The constants of the transducer are thus completely determined, and if the measurements are taken in air and in water, the radiation resistance can be separated from the frictional resistance, subject to limitations of a nature similar to those mentioned in the piezoelectric case.

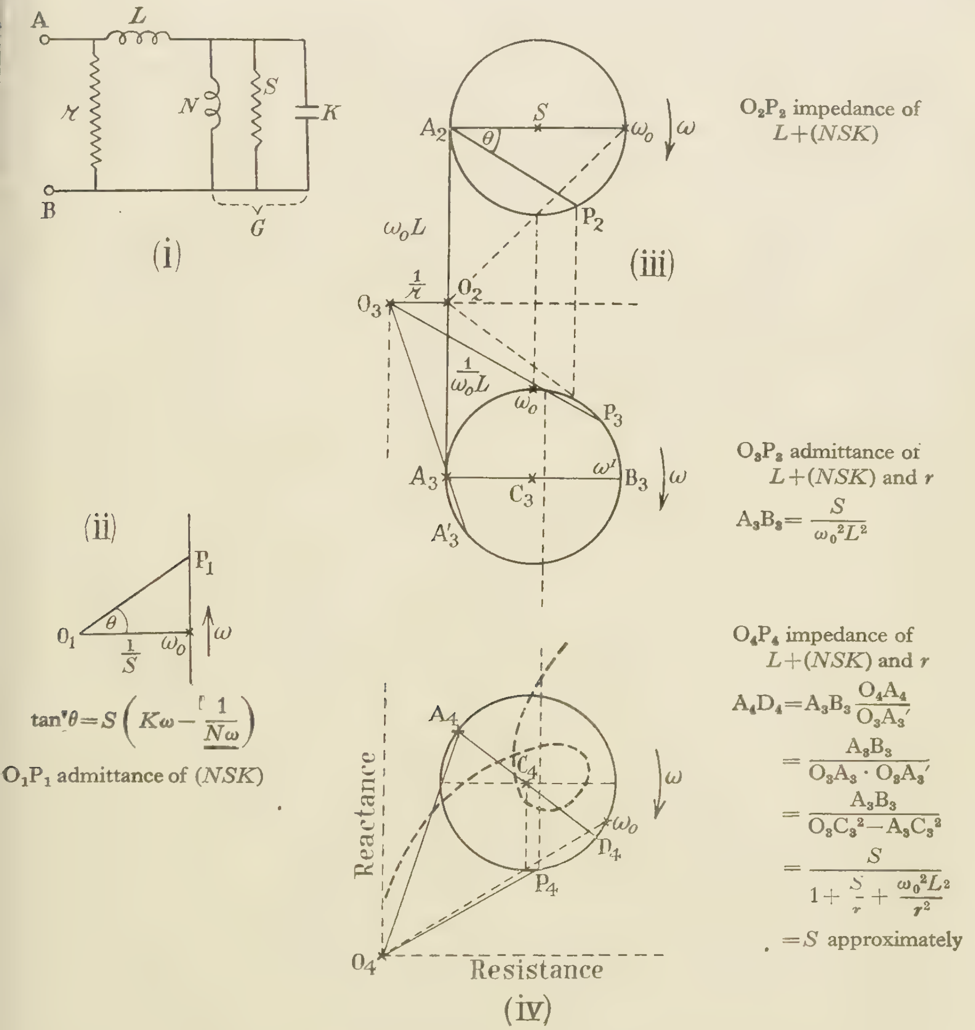

A
L
B
r
N
S
K
G
(i)
(ii)
P_1
O_1
θ
$\frac{1}{S}$
ω_0
ω
$\tan\theta = S\left(K\omega - \frac{1}{N\omega}\right)$
O_1P_1 admittance of (NSK)
A_2
S
θ
ω_0
ω
P_2
$\omega_0 L$
(iii)
O_3
$\frac{1}{r}$
O_2
$\frac{1}{\omega_0 L}$
ω_0
P_3
ω'
A_3
C_3
B_3
A'_3
O_2P_2 impedance of $L+(NSK)$
O_3P_3 admittance of $L+(NSK)$ and r
$A_3B_3 = \frac{S}{\omega_0^2 L^2}$
A_4
C_4
ω_0
D_4
P_4
O_4
Reactance
Resistance
(iv)
O_4P_4 impedance of $L+(NSK)$ and r
$A_4D_4 = A_3B_3 \frac{O_4A_4}{O_3A_3'}$
$= \frac{A_3B_3}{O_3A_3 \cdot O_3A_3'}$
$= \frac{A_3B_3}{O_3C_3^2 - A_3C_3^2}$
$= \frac{S}{1 + \frac{S}{r} + \frac{\omega_0^2 L^2}{r^2}}$
$= S$ approximately

Figure 4. Impedance diagram for magnetostriction transducer.

Even when the Q value is large, as for instance in the measurement in air, the quick method described for piezoelectric transducers, whereby all the constants can be determined from the two maxima and the corresponding frequencies, does not prove so convenient here, because the loss in the winding cannot be neglected as was the loss in the condenser, and in consequence it cannot be assumed that the circle is tangential to the impedance vector of the winding. The formulae of

§ 5 and figure 3 come out easily and simply, precisely because OA is a tangent to the circle.

For completeness, the ideal diagram of the magnetostriction transducer is given in figure 4, but in practice the reactance ωL of the winding varies considerably over the range of frequencies concerned, and cannot be regarded as constant as in the diagram. As a result of this variation the diagram is distorted into a loop, roughly indicated by the dotted curve of figure 4 (iv), and it is only after the subtraction of impedances mentioned above has been performed that a proper impedance circle is obtained.

Even after subtraction, the admittance motional circle is not always tangential to O_2A_3, as in figure 4 (iii), which is obtained under the assumption that the whole of the winding loss is due to eddy currents and, therefore, equivalent to a shunt resistance r. Allowance of a series resistance leads to a tilted admittance circle as well as a tilted impedance circle.

§ 7. ALTERNATIVE EQUIVALENT CIRCUIT

It is possible to obtain an almost identical admittance on the assumption that the motional impedance of the magnetostriction transducer is a series circuit N', S', K' connected in parallel with L and with r, as in figure 5 (i). In order to examine the extent to which the two circuits are equivalent, it is sufficient to calculate and inspect the expressions for their admittances Y, Y' for any frequency ω; since r occurs alike in both, it can be left out. Then

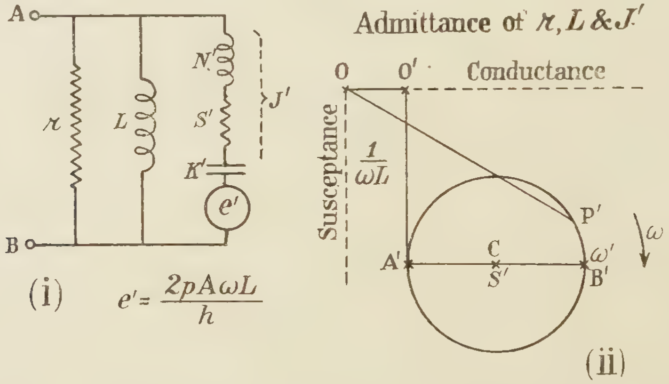


Figure 5. Alternative electrical equivalent of magnetostriction receiver.

$$\left.\begin{aligned} Y &= \frac{1}{j\omega L}\cdot\frac{1-\omega^2NK+j\omega N/S}{1+\dfrac{N}{L}-\omega^2NK+j\omega N/S},\\ Y' &= \frac{1}{j\omega L}\cdot\frac{1-\omega^2N'K'\left(1+\dfrac{L}{N'}\right)+j\omega K'S'}{1-\omega^2N'K'+j\omega K'S'}.\end{aligned}\right\}\quad\ldots\ldots(14)$$

It is not possible to choose N', S', K' so as to make Y' equal to Y at all frequencies; but in the neighbourhood of resonance the two expressions will be very nearly equal if

$$NN' = L^2,\quad SS' = \frac{L^2}{KN}\left(1+\frac{N}{L}\right),\quad\text{and}\quad \frac{K'}{K} = \frac{N^2}{L^2}\frac{1}{1+N/L}.\quad\ldots\ldots(15)$$

In particular, these values give identical admittance at the frequency for which the motional admittance is a maximum. This frequency, corresponding to B_2 in figure 4(iii), and B′ in figure 5(ii), is given by

$$\omega^2 NK = 1 + \frac{N}{L}, \quad \text{or} \quad \omega^2 = \omega_0^{\,2}\left(1 + \frac{N}{L}\right)$$

and the motional admittance is $1/S'$.

Apart, however, from the inability of the series circuit exactly to reproduce the behaviour of the transducer at all frequencies, its use as a receiver equivalent presents another difficulty: comparison of the two circuits shows that for equality of electrical output from a given sound field, an e.m.f. e' equal to $2pA\omega L\ h$ must be supposed injected into the series equivalent circuit. As this e.m.f. is not independent of frequency, the series representation, although sometimes convenient, is not strictly correct for the magnetostriction case.

§ 8. FREQUENCY OF MAXIMUM RADIATION

When the transducer is excited electrically, it is often the frequency of maximum radiation which it is desired to find, rather than that of maximum or minimum impedance. In general this frequency depends on whether the transducer is excited from constant voltage or fed with constant current, but it can always be determined by purely electrical measurements. It has been pointed out that as far as impedance or admittance diagrams are concerned, it is immaterial whether the motional component of the transducer be represented by a series network of impedance J, or a parallel network of admittance G.

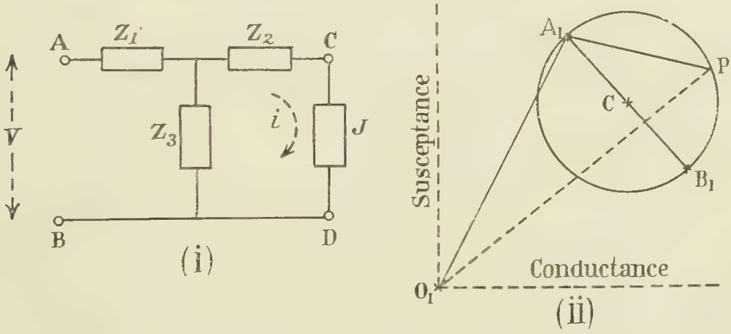


Figure 6. Admittance diagram of an electro-acoustic transducer.

To study the constant voltage case, it is convenient to use the series equivalent J. Now the whole of the purely electrical network between A and B, figure 1 or figure 5(i), can be replaced by a star network Z_1, Z_2, Z_3, connected to A, B at one end, and to J at the other, as in figure 6(i). With constant voltage applied to AB, the current or admittance O_1P into A, B will be the resultant of two vectors, the slowly varying current O_1A_1 in Z_3, figure 6(ii), and the rapidly varying current A_1P in J. If the admittance diagram is loop-shaped, the motional admittance circle can always be derived by subtraction of the slowly varying components, as explained in §§ 5 and 6; the origin A_1 of the circle, which is a point on the circumference, is located in the process. If the diameter A_1B_1 be drawn, the extremity B_1

determines the frequency at which the current i in J is a maximum. Since the radiated power is Si^2, this frequency is also the frequency of maximum radiation for constant applied voltage.

The constant-current case can be studied by transforming the purely electrical network between A, B, figure 2, into a Π network Z_1, Z_2, Z_3 connected to A, B at one end, and to the parallel equivalent admittance G at the other, as in figure 7 (i). With constant current fed into A, B the voltage or impedance O_2P across AB will be the resultant of two vectors, the slowly varying voltage O_2A_2 across Z_2, figure 7 (ii), and the rapidly varying voltage A_2P across G. If the impedance diagram is loop-shaped, the motional impedance circle can again be derived by subtraction of the slowly varying components, and the extremity B_2 of the diameter through the origin A_2 determines the frequency for which the voltage v across G is greatest. Since the radiated power is v^2/S, this frequency is also the frequency of maximum radiation for constant current fed to the transducer.

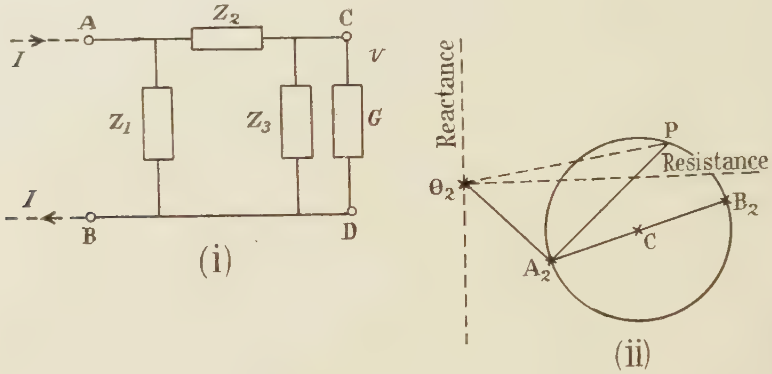


Figure 7. Impedance diagram of an electro-acoustic transducer.

In general, a transducer is neither excited from constant voltage nor fed with constant current, but is coupled to a generator of constant e.m.f. E, having a certain internal impedance, say Z_e. But the complete system can be represented by a constant voltage E applied to the impedance of the transducer and the impedance Z_e in series. Thus this general case reduces to that of constant applied voltage, and can be solved by production of the admittance diagram, figure 6.

In practice, condensers or coils are added to the circuit so as to "tune" it, i.e. to arrange that its admittance at the frequency of maximum radiation would be a maximum if the transducer were clamped. The condensers or coils also serve to "match" the transducer to the generator, i.e. to ensure that half the power generated is radiated. On account of the "tuning", the electrical impedances vary with frequency in the neighbourhood of resonance nearly as much as the motional impedance, and the admittance diagram cannot be derived in the simple way illustrated in figures 3 and 4. But in this case a readier method of determining the frequency of maximum radiation and at the same time of ensuring correct tuning is to plot current against frequency, when the "M" or double-hump curve characteristic of coupled circuits is obtained. Trial adjustment of the tuning condenser, without entailing much labour, secures a symmetrical curve, the minimum giving the frequency of maximum radiation.

ACKNOWLEDGMENT

Thanks are due to the Chief of the Royal Naval Scientific Service for permission to publish this paper.

APPENDIX

Rapid analysis of piezoelectric transducer

When the Q value is high, i.e. when $SK\omega_0$ is small, the ends of the admittance vectors lie almost exactly on a circle. In figure 3 (ii), AB is proportional to $1/S$, where for brevity S is written for the sum of the internal and radiation resistances, and OA is proportional to $C\omega_0$. The approximate values are

$$S=\frac{v}{\mathrm{AB}}=\frac{v}{i_1-i_2}, \qquad \ldots\ldots(16)$$

$$C=\frac{\mathrm{AO}}{v\omega_0}=\frac{\sqrt{\mathrm{OC}\,.\,\mathrm{OD}}}{v\omega_0}=\frac{\sqrt{i_1 i_2}}{v\omega_0}. \qquad \ldots\ldots(17)$$

The frequencies ω_1, ω_2 are related to the transducer constants by the formulae

$$\tan \mathrm{DAB}=\frac{\dfrac{1}{\omega_1 K}-\omega_1 N}{S}=\frac{\omega_0^2-\omega_1^2}{\omega_0\omega_1}Q, \qquad \ldots\ldots(18)$$

$$\tan \mathrm{CAB}=\frac{\omega_2 N-\dfrac{1}{\omega_2 K}}{S}=\frac{\omega_2^2-\omega_0^2}{\omega_0\omega_2}Q. \qquad \ldots\ldots(19)$$

But also

$$\tan \mathrm{DAB}=\tan \tfrac{1}{2}\mathrm{AHO}=\sqrt{\frac{1-\cos \mathrm{AHO}}{1+\cos \mathrm{AHO}}}=\sqrt{\frac{\mathrm{OH}-\mathrm{AH}}{\mathrm{OH}+\mathrm{AH}}}=\sqrt{\frac{\mathrm{OC}}{\mathrm{OD}}}=\sqrt{\frac{i_2}{i_1}}. \qquad \ldots\ldots(20)$$

$$\tan \mathrm{CAB}=\tan \tfrac{1}{2}\mathrm{AHD}=\sqrt{\frac{1-\cos \mathrm{AHD}}{1+\cos \mathrm{AHD}}}=\sqrt{\frac{\mathrm{OH}+\mathrm{AH}}{\mathrm{OH}-\mathrm{AH}}}=\sqrt{\frac{\mathrm{OD}}{\mathrm{OC}}}=\sqrt{\frac{i_1}{i_2}}. \qquad \ldots\ldots(21)$$

Insertion of these values in (18), (19) and division gives after some rearrangement

$$\omega_0^2=\omega_1\omega_2\frac{\omega_1 i_1+\omega_2 i_2}{\omega_2 i_1+\omega_1 i_2}=\omega_1\omega_2(1-\alpha) \qquad \ldots\ldots(22)$$

where

$$\alpha\equiv\frac{(\omega_2-\omega_1)(i_1-i_2)}{\omega_2 i_1+\omega_1 i_2}$$

is small.

Thus to a first approximation ω_0 is equal to $\sqrt{\omega_1\omega_2}$.

The value of Q can be found by adding equations (20) and (21) and equating to the sum of (18) and (19) which, after some reduction, gives

$$Q(\omega_0^2+\omega_1\omega_2)(\omega_2-\omega_1)=\omega_0\omega_1\omega_2\left(\sqrt{\frac{i_1}{i_2}}+\sqrt{\frac{i_2}{i_1}}\right)$$

or

$$Q=\frac{\sqrt{\frac{i_1}{i_2}}+\sqrt{\frac{i_2}{i_1}}}{\omega_2-\omega_1}\cdot\frac{\sqrt{\omega_1\omega_2}}{\frac{\omega_0}{\sqrt{\omega_1\omega_2}}+\frac{\sqrt{\omega_1\omega_2}}{\omega_0}}.$$

By (22), the denominator of the second factor becomes

$$\begin{aligned}\frac{\omega_0}{\sqrt{\omega_1\omega_2}}+\frac{\sqrt{\omega_1\omega_2}}{\omega_0}&=(1-\alpha)^{\frac{1}{2}}+(1-\alpha)^{-\frac{1}{2}}\\&=2(1+\tfrac{1}{8}\alpha^2),\end{aligned}$$

which is in general very nearly equal to 2. Thus the formula becomes

$$Q=\frac{\sqrt{\omega_1\omega_2}}{\omega_2-\omega_1}\cdot\frac{1}{2}\left(\sqrt{\frac{i_1}{i_2}}+\sqrt{\frac{i_2}{i_1}}\right). \qquad \ldots\ldots(23)$$

Since Q is equal to $SK\omega_0$, K is obtained from (23) and (16).

LIGHT ABSORPTION AND SELECTIVE PHOTO-EFFECT IN ADSORBED LAYERS

By H. FRÖHLICH and R. A. SACK,*
H. H. Wills Physical Laboratory, University of Bristol

* Now at Department of Applied Mathematics, University of Liverpool

MS. received 1 *July* 1946

ABSTRACT. An estimate is often required of the absorption of light by a single layer of atoms (in particular alkali metal atoms) adsorbed on a non-metallic surface. Such an absorption is accompanied in many cases by a selective photoelectric effect at comparatively long wave-lengths. In the present paper the probability is calculated of a light quantum being absorbed by such a layer, and a mechanism suggested for the emission of the photo-electrons.

§ 1. ABSORPTION OF INCIDENT LIGHT

THE absorption of a light quantum leads to the transition of an electron in an adsorbed atom from the ground state to a higher level. As with the absorption of light by atoms dissolved in solids, we should expect a considerable broadening of the absorption line, i.e. practically an absorption band about the frequency ν_0, where $h\nu_0$ is the energy difference between the ground level and the excited level of the electron. For reasons shown by Mott and Gurney (1940, p. 116) the width of the absorption band is proportional to $\sqrt{T}$, where T is the absolute temperature.

To find a general expression for the absorption we bear in mind that, as far as the transition from the ground state to the excited state is concerned, a number N of adsorbed atoms behave like Nf classical harmonic oscillators of electronic harge and mass, where f denotes the oscillator strength connected with the ransition. Let

$$E=E_0\cos 2\pi\nu t \qquad \ldots\ldots(1)$$

e the homogeneous electric field acting on the adsorbed layer; the polarization nduced in it will be of the form

$$P=P_1\cos 2\pi\nu t+P_2\sin 2\pi\nu t. \qquad \ldots\ldots(2)$$

t is useful to define a complex dielectric constant

$$\epsilon=\epsilon_1-i\epsilon_2 \qquad \ldots\ldots(3)$$

n such a way that the electric displacement

$$D=E+4\pi P \qquad \ldots\ldots(4)$$

s given by the real part of $\epsilon E_0\, e^{2\pi i\nu t}$. Then

$$4\pi P_1=(\epsilon_1-1)E_0; \qquad 4\pi P_2=\epsilon_2 E_0. \qquad \ldots\ldots(5)$$

If we assume that the magnetic field H does not act upon the electron, the nergy absorbed by the adsorbed atoms per unit time and volume is, according o Maxwell's equations, given by

$$\bar{\dot{u}}=\frac{1}{4\pi}\overline{E\dot{D}}, \qquad \ldots\ldots(6)$$

here the bars indicate averages over one period.

If the number of atoms per cm.^{2} in the surface layer is denoted by z, and their umber per unit volume by N, the rate of absorption per unit surface is given by z/N. Substituting into (6) from (2) and (4) and making use of (5) we find

$$\frac{z}{N}\bar{\dot{u}}=\frac{z}{N}\frac{\epsilon_2\nu E_0{}^2}{4}. \qquad \ldots\ldots(7)$$

Thus the probability p for the absorption of light is given by

$$p=\frac{z}{N}\frac{\bar{\dot{u}}}{I}=\frac{z}{N}\frac{\epsilon_2\nu E_0{}^2}{4I}, \qquad \ldots\ldots(8)$$

I is the intensity of the incident light. Since the average values of electric nd magnetic energies in a light wave are equal, we have

$$I=\frac{\boldsymbol{c}}{8\pi}(\overline{E^2}+\overline{H^2})=\frac{\boldsymbol{c}\overline{E^2}}{4\pi}=\frac{\boldsymbol{c}E_0{}^2}{8\pi}. \qquad \ldots\ldots(9)$$

nsertion of this into (8) leads to

$$p=\frac{z}{N}\frac{2\pi\nu\epsilon_2}{\boldsymbol{c}}. \qquad \ldots\ldots(10)$$

ere it is assumed that the velocity of light in the surface layer is approximately qual to the velocity $\boldsymbol{c}$ in vacuum, and that the local electric field acting on an lectron is equal to the field strength in vacuum. These assumptions are equialent to the conditions

$$\epsilon_1-1\ll 1, \qquad \epsilon_2\ll 1, \qquad \ldots\ldots(11)$$

that the interaction between the induced polarizations of different atoms can e neglected. It also excludes the possibility of appreciable reflexion from a upporting layer. Such terms would not, however, change the order of magnitude (10).

The shape and magnitude of the absorption band is determined by ϵ_2. From very general arguments by S. Whitehead (1946) and B. Gross (1941) it follows, with the use of (11), that

$$\Delta\epsilon = \frac{2}{\pi}\int_0^\infty \epsilon_2(\nu)\,\frac{d\nu}{\nu}, \qquad \ldots\ldots(12)$$

where $\Delta\epsilon$ is the contribution of the oscillators to the static dielectric constant. If we assume that the absorption band is centred about the proper frequency ν_0 of the oscillators and is of a mean width $\Delta\nu$, the average value $\bar{\epsilon}_2$ is found from (12):

$$\Delta\epsilon \simeq \frac{2}{\pi}\,\frac{\Delta\nu}{\nu_0}\,\bar{\epsilon}_2, \qquad \ldots\ldots(13)$$

so that the average probability of absorption within the absorption range will be given, according to (10), by

$$\bar{p} = \frac{z}{N}\,\frac{\pi^2\nu_0{}^2\Delta\epsilon}{\boldsymbol{c}\Delta\nu}. \qquad \ldots\ldots(14)$$

Now the restoring force of an elastically bound electron is $4\pi^2 m\nu_0{}^2\Delta x$, and thus the average displacement in a constant field E_0 is $\Delta x = eE_0\;4\pi^2 m\nu_0{}^2$. Hence, since there are Nf such oscillators per unit volume,

$$\Delta\epsilon = \frac{4\pi e\Delta x Nf}{E_0} = \frac{e^2 Nf}{\pi\nu_0{}^2 m}. \qquad \ldots\ldots(15)$$

Inserting this into (14), we obtain

$$\bar{p} = \frac{\pi e^2 zf}{m\boldsymbol{c}\Delta\nu} = \frac{\pi e^2\lambda_0 zf}{m\boldsymbol{c}^2}\,\frac{\nu_0}{\Delta\nu}, \qquad \ldots\ldots(16)$$

where $\lambda_0 = \boldsymbol{c}/\nu_0$.

Thus for an absorption band in the visible ($\lambda_0 \simeq 5\times10^{-5}$ cm.) with a large oscillator strength ($f \simeq 1$) and a broadening $\Delta\nu/\nu_0 \simeq 1/2$, we obtain for a density of $z = 10^{15}$ atoms per unit surface ($e^2/m^2\boldsymbol{c} = 3\times10^{-13}$ cm.)

$$\bar{p} = 9\times10^{-2},$$

i.e. absorption of about 10%.

§ 2. PHOTO-EMISSION

Emission of photoelectrons from adsorbed atoms can be obtained by the impact of light quanta with an energy less than the ionization energy of the free atom. The reason for this is that part of the ionization energy may be provided by the difference in the energies of adsorption of the neutral atom and the positive ion. Photo-emission may thus be considered as a two-stage process: (i) absorption of a quantum $\boldsymbol{h}\nu$ in the absorption band centred around $\boldsymbol{h}\nu_0$ leading to a transition of the electron from the ground state to an excited level; (ii) emission of the electron from the excited state.* Two mechanisms can be suggested for the second step. Either the energy difference Δ between the adsorption energies of the excited neutral atom and the positive ion is larger than the energy α required for the emission of an electron from the excited atom (and in this case electrons will be emitted with an average kinetic energy $\alpha - \Delta$), or else α is slightly smaller than Δ;

---

* A similar two-stage process has been suggested by Ryzhanov (1939); as the second step he assumes a transition of the excited metal electron into a lower state with a simultaneous excitation or emission of an electron in the supporting semi-conductor.

then the missing energy may be provided by the thermal energy of the supporting solid. This latter possibility is similar to the case of the internal photo-effect (cf. Mott and Gurney, 1940, p. 134) and, as in that case, a sudden drop in the photo-yield is to be expected as the temperature is lowered below a critical value. For this second mechanism the kinetic energy of the emitted electrons should be of the order of the thermal energy. For both mechanisms suggested, the selective photo-effect should be centred about a frequency corresponding to a strong absorption line of the free atom. Actually the maximum photo-yield of Cs atoms adsorbed on a layer of Cs_2O lies at 8000 A. (cf. Kluge, 1933; de Boer, 1935, p. 327), whereas the resonance line for the free Cs atoms form a doublet at about 8500 A. and 8900 A. For Rb on Rb_2O the corresponding figures are: 6500–6800 A. for the yield maximum and 7900 A. for the resonance line. This small shift of the maximum of the selective photo-effect can be expected in view of the different adsorption energies of the atoms in the ground state and excited state. For potassium the agreement is less satisfactory; the resonance line is at approximately 7700 A., whereas Kluge (1933) finds the maximum photo-yield at 4600–5200 A.; this peak, however, is not very pronounced, and its position cannot be accurately determined because of the proximity of a much stronger maximum at 4100 A., which is due to the supporting layer.

Another conclusion arising out of the present theory is that the spectral width of the emission band should increase proportional to $\sqrt{T}$, similar to the increase in the width of the absorption band in solids. Thus, the higher the temperature, the further should the photoelectric threshold be shifted towards longer wavelengths. No experimental evidence on this point is available at present.

REFERENCES

de Boer, J. H., 1935. *Electron Emission and Adsorption Phenomena* (Cambridge: The University Press).
Gross, B., 1941. *Phys. Rev.*, **59**, 748.
Kluge, W., 1933. *Phys. Z.*, **34**, 125.
Mott, N. F. and Gurney, R. W., 1940. *Electronic Processes in Ionic Crystals* (Oxford: The Clarendon Press).
Ryzhanov, S., 1939. *J. Exp. Theor. Phys., U.S.S.R.*, **9**, 38.
Whitehead, S., 1946. *Trans. Faraday Soc.* (in press).

CORRIGENDUM

"The Propagation of Supersonics in Capillary Tubes", by J. May (*Proc. Phys. Soc.*, **50**, 558 (1938)).

The figures in the seventh column of the table, headed "Amplitude absorption coefficient, Practical value", should be divided by 2.

(The author wishes to thank Mr. J. E. Drummond for pointing out an arithmetical error.)

FLUCTUATIONS IN STREAMS OF THERMAL RADIATION

By W. B. LEWIS, F.R.S.,
Telecommunications Research Establishment, Ministry of Supply

MS. received 13 *July* 1946

ABSTRACT. An ultimate limit to the sensitivity of a detector of thermal radiation is set by the inherent fluctuations in the stream of radiation. This paper presents the relevant quantitative relations to enable these fluctuations to be calculated in practical cases. The result is presented as the minimum detectable periodic modulation of the power of a stream for general conditions under which the detector may receive streams of radiation from surroundings at different temperatures and thermal equilibrium is not necessarily established. The modification necessary when the spectral frequency-distribution differs from that of full black-body radiation is also presented.

Limiting factors other than radiation fluctuations are not discussed.

§1. INTRODUCTION

CONSIDERATION of the fluctuations of radiation in a defined volume has played an important part in the history of the theory of radiation. It was shown by Einstein (1909) that the fluctuations could be represented as the sum of two terms, one which would be accounted for by classical wave-theory and the other having no interpretation on this theory but representing exactly the fluctuations expected on an extreme light-quantum view in which the quanta are assumed to behave as "classical" particles (e.g. Fowler, 1936, p. 765).

It is known, however, that an attempt to relate Planck's radiation law to such a view of a light-quantum or photon "gas" fails (e.g. Born, 1935, p. 215). The discrepancy has been traced to unwarranted assumptions such as the identity of individual quanta in the application of statistics, and when statistics omitting such assumptions are applied (Einstein–Bose statistics) to the photon gas, agreement with Planck's formula is achieved (see Born, *loc. cit.* p. 223). Moreover, deducing the fluctuations of radiation in a defined volume by applying Einstein–Bose statistics to a photon gas, the correct relation established by Einstein (1909) is obtained (Fürth, 1928; Fowler, *loc. cit.* p. 765).

It may therefore be assumed that full reliance may be placed on deductions of fluctuations in streams of radiation made by applying correct statistics to radiation treated as a photon gas.

It appears that although so much attention has been focused on radiation fluctuations, an explicit expression directly applicable to fluctuations limiting radiation pyrometry has not previously been published. It is the aim of this paper to set the theoretical knowledge of radiation fluctuations in the form most suited to the applied science of electronics by which such fluctuations may be observed.

In the field of electronics there is much experience of fluctuations or "noise" and the relevant theory is well developed; it will therefore be convenient to draw a parallel with the theory of the shot effect in presenting the results.

A radiation pyrometer may consist of a small thermal detecting element on to which the radiation is focused from a mirror. The equilibrium temperature of the detector will be determined by the balancing of the streams of radiation from the source via the mirror and from other surroundings against the heat conducted by the supports. It is assumed that the detector is vacuum mounted. Fluctuations of the detector temperature will arise from the fluctuations of the heat flow along these paths.

In critical comment it may be noted that as usual in practical considerations of fluctuations the assumption is being made that the whole apparatus may be divided into two parts. Discussion of fluctuations is related to one part while the other is assumed to obey the ordinary laws of large scale continuous physics (Fowler, *loc. cit.* p. 788). Here the division is being made at the boundary of the thermal detecting element. In practice it will almost inevitably be necessary to take account of fluctuations of voltage or current flow in the element, and also in the amplifier system used for detection. It is therefore desirable to present the results of this consideration of radiation fluctuations in such a form that these other fluctuations arising in the observing system may readily be incorporated.

To narrow the problem to determining the fluctuations of a stream of radiation, suppose the detector is idealized as an infinitesimal speck of matter enclosed in a small hollow sphere of volume v with diffusing perfectly reflecting internal walls in which there is a very small aperture of area a. The detector is thus an ideal black body of surface area a with a thermal capacity which is that of the enclosed radiation.

§2. FLUCTUATIONS IN A DEFINED VOLUME

The fluctuation of radiation energy in such an enclosure but without the speck of matter was the problem studied by Einstein in 1909. Writing η as the radiation energy of frequency ν to $\nu+d\nu$ in v, at any given instant η will differ from the mean η_0 by a small quantity ϵ so that $\eta=\eta_0+\epsilon$.

It is necessary to assume that $v \gg \lambda^3$ so that the energy is statistically significant, where λ is the wave-length corresponding to ν.

Using the general statistical relations between entropy and probability, and determining the entropy of radiation over a frequency range ν to $\nu+d\nu$ from Planck's radiation formula, Einstein arrived at the result

$$\overline{\epsilon^2}=\eta_0 \boldsymbol{h}\nu+\frac{\boldsymbol{c}^3}{8\pi\nu^2 d\nu}\frac{\eta_0{}^2}{v}. \qquad \ldots\ldots(1)$$

Regarding the radiation as a light-quantum gas and writing n = number of quanta in volume v, we have

$$\eta_0=n\boldsymbol{h}\nu, \qquad \ldots\ldots(2)$$

so the first term of equation (1) is $n(\boldsymbol{h}\nu)^2$. This is the fluctuation to be expected on the basis of classical statistics, but it is known that such statistics are not applicable to light-quanta in this way and Einstein Bose statistics should be used instead.

From the theory of fluctuations in Einstein–Bose statistics, it is known that where in classical statistics the mean-square fluctuation of the number of particles $\Delta n^2 = n$, in the Einstein–Bose statistics the mean-square fluctuation

$$\Delta n^2 = n + \frac{n^2}{N}, \qquad \ldots\ldots(3)$$

where N is the total number of independent cells which can be occupied by the particles, in this case the total number of independent standing wave vibrations in volume v.

This number is known from fundamental electro-magnetic wave theory (see e.g. Roberts, 1928, p. 398) to be

$$\frac{8\pi\nu^2 d\nu}{\mathbf{c}^3} \cdot v,$$

so we have

$$\overline{\epsilon^2} = (\boldsymbol{h}\nu)^2 \Delta n^2 = (\text{from eqn. (3)})\ n(\boldsymbol{h}\nu)^2 + \frac{n^2(\boldsymbol{h}\nu)^2\mathbf{c}^3}{8\pi\nu^2 d\nu} \cdot \frac{1}{v} \qquad \ldots\ldots(4)$$

which by equation (2) may be seen to be identical with equation (1).

Noting that from Planck's radiation formula

$$\eta_0 = \frac{8\pi\nu^2 d\nu}{\mathbf{c}^3} \cdot \frac{\boldsymbol{h}\nu}{e^{\boldsymbol{h}\nu/\boldsymbol{k}T} - 1} \cdot v\ ; \qquad \ldots\ldots(5)$$

the second term of equation (1) may be recast as

$$\frac{\eta_0 \boldsymbol{h}\nu}{e^{\boldsymbol{h}\nu/\boldsymbol{k}T} - 1}\ ;$$

the relative magnitude of the two terms may be appreciated by rewriting equation (1) as

$$\overline{\epsilon^2} = \eta_0 \boldsymbol{h}\nu \left[1 + \frac{1}{e^{\boldsymbol{h}\nu/\boldsymbol{k}T} - 1}\right]. \qquad \ldots\ldots(6)$$

The second term is relatively large only when $\boldsymbol{h}\nu \ll \boldsymbol{k}T$, in which limit the expression tends to $\overline{\epsilon^2} = \eta_0 \boldsymbol{k}T$.

§ 3. INTEGRATION OF FLUCTUATIONS OVER THE SPECTRUM

It is of interest to follow out the integration of $\overline{\epsilon^2}$ over all values of ν, because in practice it may be desired to appreciate the effect of absorption bands in parts of the spectrum.

By substituting the value of η_0 given by equation (5) in equation (6) we obtain

$$\overline{\epsilon^2} = \frac{8\pi\nu^2 d\nu}{\mathbf{c}^3} \cdot \frac{(\boldsymbol{h}\nu)^2 e^{\boldsymbol{h}\nu/\boldsymbol{k}T}}{(e^{\boldsymbol{h}\nu/\boldsymbol{k}T} - 1)^2} \cdot v. \qquad \ldots\ldots(7)$$

Writing $\xi = \boldsymbol{h}\nu/\boldsymbol{k}T$, equation (7) becomes

$$\overline{\epsilon^2} = (\boldsymbol{k}T)^5 \frac{8\pi v}{\mathbf{c}^3 \boldsymbol{h}^3} \cdot \frac{\xi^4 e^\xi}{(e^\xi - 1)^2} \cdot d\xi.$$

The total energy fluctuation is

$$\overline{\Delta E_v{}^2} = (\boldsymbol{k}T)^5 \frac{8\pi v}{\mathbf{c}^3 \boldsymbol{h}^3} \int_0^\infty \frac{\xi^4 e^\xi}{(e^\xi - 1)^2} \cdot d\xi.$$

The function $\frac{\xi^4 e^\xi}{(e^\xi-1)^2}$ is shown in figure 1.

Write
$$y=\frac{\xi^4}{e^\xi-1};$$

then
$$\frac{dy}{d\xi}=\frac{4\xi^3}{e^\xi-1}-\frac{\xi^4 e^\xi}{(e^\xi-1)^2}.$$

$$\therefore \quad \overline{\Delta E_v^{\,2}}=(\boldsymbol{k}T)^5\frac{8\pi v}{\mathbf{c}^3\boldsymbol{h}^3}\left[\int_0^\infty-\frac{dy}{d\xi}\,.\,d\xi+\int_0^\infty\frac{4\xi^3}{e^\xi-1}\,.\,d\xi\right]$$

$$=(\boldsymbol{k}T)^5\frac{8\pi v}{\mathbf{c}^3\boldsymbol{h}^3}\left\{\left[\frac{-\xi^4}{e^\xi-1}\right]_0^\infty+\frac{4\pi^4}{15}\right\}.$$

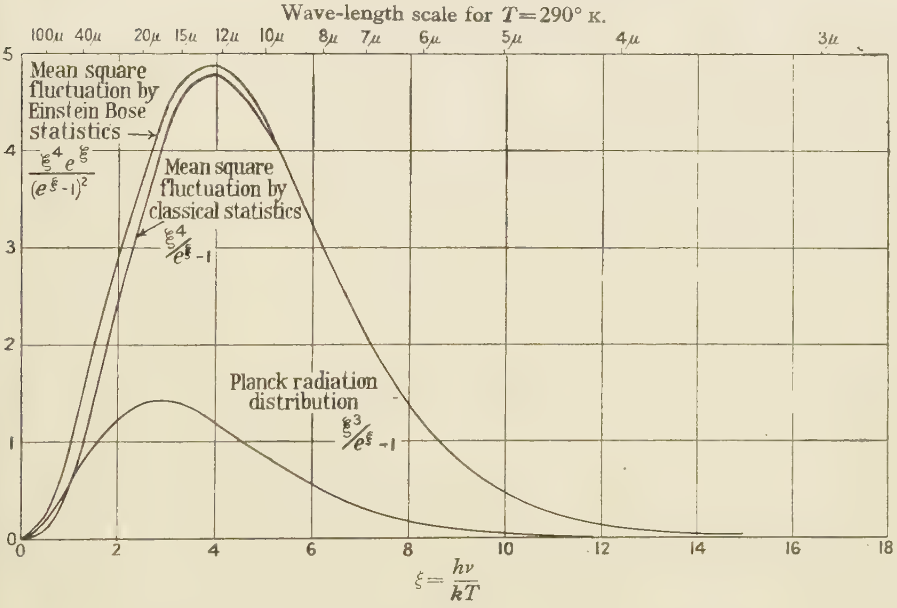


Figure 1.

The last term occurs in the integration of Planck's formula to relate it to Stefan's Law (see Wilson, 1944, p. 84), and since

$$\left[\frac{-\xi^4}{e^\xi-1}\right]_0^\infty=0,$$

we have

$$\overline{\Delta E_v^{\,2}}=\frac{(2\pi\boldsymbol{k}T)^5}{15\mathbf{c}^3\boldsymbol{h}^3}\,.\,v. \qquad \ldots\ldots(8)$$

From standard works (e.g. Roberts, 1928, p. 369) we have $E_T=\frac{4\sigma T^4}{\mathbf{c}}$ where $\sigma=$ Stefan's constant $=\frac{2\pi^5\boldsymbol{k}^4}{15\mathbf{c}^2\boldsymbol{h}^3}$ from the integration of Planck's formula, and E_T is the energy density of full black-body radiation of temperature T.

Equation (8) becomes, on substituting,

$$\overline{\Delta E_v^{\,2}}=4E_T\boldsymbol{k}Tv, \qquad \ldots\ldots(9)$$

a relation which may be obtained directly from the general relation

$$\Delta E^2 = \boldsymbol{k}T^2\frac{dE}{dT} \qquad \ldots\ldots(10)$$

(Fowler, *loc. cit.*, equation 2156, p. 764), noting that

$$\frac{dE_T}{dT} = \frac{16\sigma T^3}{\mathbf{c}} = \frac{4E_T}{T}.$$

It may be of interest to note that if the second term of equation (1) were omitted we should have obtained instead of equation (9) $\overline{\Delta E_v^2} = 3{\cdot}83 E_T \boldsymbol{k} T v$, and as already noted, the contribution of the second term is mainly in the low frequency or long wave-length region.

§4. FLUCTUATIONS OF RADIATION ENERGY CROSSING AN AREA

Regarding radiation as a light-quantum gas consisting of quanta of energy $\boldsymbol{h}\nu$ moving in all directions with velocity $\mathbf{c}$, then the number crossing an area a per second would be $\frac{1}{4}n_1\mathbf{c}a$ where n_1 is the number per unit volume. Applying Einstein-Bose statistics, the mean-square fluctuation $\overline{\Delta E_a^2}$ of the radiation of frequency ν crossing an area a would be

$$\tfrac{1}{4}n_1\mathbf{c}a(\boldsymbol{h}\nu)^2\left[1+\frac{n_1}{N}\right].$$

Noting that $\eta_0 = n_1\boldsymbol{h}\nu v$, this may be written

$$\overline{\Delta E_{av}^2} = \frac{\mathbf{c}a}{4v}\,\eta_0\boldsymbol{h}\nu\left[1+\frac{1}{e^{\boldsymbol{h}\nu/\boldsymbol{k}T}-1}\right]$$

$$= \frac{\mathbf{c}a}{4v}\,\overline{\epsilon^2} \quad \text{per second (from equation (6)).} \qquad \ldots\ldots(11)$$

The total energy fluctuation may be obtained by integrating (11) over all frequencies. By comparison with equation (8) this may be written down at once as

$$\Delta E_a^{\,2} = \frac{\mathbf{c}a}{4}\cdot\frac{(2\pi\boldsymbol{k}T)^5}{15\mathbf{c}^3\boldsymbol{h}^3} \quad \text{or by equation (9)} = E_T\boldsymbol{k}T\mathbf{c}a. \qquad \ldots\ldots(12)$$

§5. TIME VARIATION OF FLUCTUATIONS IN A RADIATION STREAM

Equation (12) is a relation closely parallel with the basic idea of the shot-effect fluctuations of the saturation current in a thermionic diode. If λ is the mean rate of arrival of electrons at the anode, the mean-square fluctuation of the charge reaching the anode per second is $\overline{q^2} = e^2\lambda$ where e is the electronic charge. But $e\lambda = I_0$, the mean current, so that

$$\overline{q^2} = eI_0. \qquad \ldots\ldots(13)$$

In practical considerations of shot noise in circuits it is convenient to use a related equation

$$\overline{i_f^2} = 2eI_0\Delta f \qquad \ldots\ldots(14)$$

where $\overline{i_f^2}$ is the mean square current fluctuation over the frequency range f to

$f\Delta + f$. The transition from equation (13) to equation (14) has been established by a necessarily elaborate mathematical argument, and it may be taken over for the thermal case. When applied to equation (12) we obtain

$$\overline{\Delta W^2_{af}} = 2\boldsymbol{k}TE_T\boldsymbol{c}a\Delta f \qquad \ldots\ldots(15)$$

where W_a is the rate of flow of energy or power crossing a.

It is to be understood that this may be applied to determine the frequency spectrum of the fluctuations generated in a body exposed to this stream of radiation.

Moreover, if the body is black and at temperature T and therefore radiating an equivalent stream to balance the received radiation, then the effective mean square fluctuation of energy flow from the body over the frequency range Δf is doubled, viz.,

$$\overline{\Delta W^2_{bf}} = 4\boldsymbol{k}TE_T\boldsymbol{c}a\Delta f, \qquad \ldots\ldots(16)$$

or, since $E_T\boldsymbol{c} = 4\sigma T^4$, where $\sigma =$ Stefan's constant,

$$\overline{\Delta W^2_{bf}} = 16\sigma\boldsymbol{k}T^5a\Delta f$$

or

$$\sqrt{\overline{\Delta W^2_{bf}}} = 4T^2\sqrt{a\boldsymbol{k}T\Delta f}, \qquad \ldots\ldots(17)$$

a relation established by Daunt (1945)* for the minimum power detectable as a sinusoidal variation of frequency f by an isolated body connected to its surroundings only by radiation.

§6. RADIATION FLUCTUATIONS IN PRACTICE

The above argument by which this relation has been reached indicates how this minimum detectable power may be determined when the detector receives streams of energy from surroundings at differing temperatures and thermal equilibrium is not necessarily assumed. Using equation (15) for each stream, the total mean-square fluctuation may be determined by summing these contributions for all the streams.

It may be noted that in terms of σ equation (15) becomes

$$\overline{\Delta W^2_{af}} = 8\sigma\boldsymbol{k}T^5a\Delta f. \qquad \ldots\ldots(18)$$

An object at temperature T_1 subtending a small solid angle α in a direction making an angle θ with the normal to the plane of a will contribute a term

$$\overline{\Delta W^2_{af_1}} = 32\sigma\boldsymbol{k}T_1{}^5\,a\cos\theta\Delta f\,.\,\frac{\alpha}{4\pi}. \qquad \ldots\ldots(19)$$

The magnitude of this contribution may be assessed, taking $\sigma = 5{\cdot}65 \times 10^{-5}$ erg. cm.$^{-2}$ sec.$^{-1}$ deg.$^{-4}$, $\boldsymbol{k} = 1{\cdot}38 \times 10^{-16}$ erg. deg.$^{-1}$, $a = 1$ mm., $\Delta f = 1$ c./s., and writing $T_0 = 290°$ K., we have

$$\overline{\Delta W^2_{af_1}} = 0{\cdot}97 \times 10^{-24}\,\alpha\cos\theta(T_1/T_0)^5 \text{ watt}^2.$$

The minimum detectable power as defined by Daunt (1945) will be

$$\sqrt{\sum_{n=1}^{n=n}\overline{W^2_{af_n}}} = 0{\cdot}98 \times 10^{-12}\left\{\pi\left(\frac{T}{T_0}\right)^5 + \sum_{n=1}^{n=n}\alpha_n\cos\theta_n\left(\frac{T_n}{T_0}\right)^5\right\}^{\frac{1}{2}} \text{ watts}$$

where the summation is carried out over all objects in view, and T is the temperature of the detector. This assumes that the detector and all objects are perfectly black. Modifications for grey and reflecting objects will be obvious, but it may be noted that if the object reflects radiation back to the detector, neither the outgoing nor the reflected stream contributes to the fluctuations if the distance of the reflector is $\ll \boldsymbol{c}/\Delta f$.

---

* Also independently by Milatz (1943).

It may be noted for comparison that the total radiation crossing a, namely $\sigma a T_0^4 = 4 \times 10^{-4}$ watts, while for $T = T_0$ from equation (17) the minimum detectable power for $\Delta f = 1$ c./s. is $2{\cdot}46 \times 10^{-12}$ watts. On the view presented in this paper, this minimum detectable power is determined by the statistical variation in the number of quanta crossing a per second. The average energy of the quanta is about $4\boldsymbol{k}T$ and the number n_2 crossing a per second is about 10^{16}. The fluctuation of the number is about $\sqrt{n_2} \approx 10^8$, so the minimum detectable power is about 10^{-8} of the total flow of energy in this example.

If the radiation in any stream is not full but the spectral distribution is known to be $f(\nu)E$ (where E represents the spectral distribution for full black-body radiation) the resultant fluctuations may be calculated by the following relation (from equations (7), (11), (15) and (19)):

$$\overline{\Delta W^2_{af_1}} = \frac{4(\boldsymbol{k}T_1)^5}{\boldsymbol{c}^2\boldsymbol{h}^3} \,.\, a\cos\theta \,.\, \Delta f \,.\, \alpha \int_0^\infty \frac{(\boldsymbol{h}\nu/\boldsymbol{k}T_1)^4 e^{\boldsymbol{h}\nu/\boldsymbol{k}T_1}}{(e^{\boldsymbol{h}\nu/\boldsymbol{k}T_1} - 1)^2} \,.\, f(\nu)\, \frac{\boldsymbol{h}d\nu}{\boldsymbol{k}T_1} \qquad \ldots\ldots(20)$$

In this expression the integral is a numerical term which, if $f(\nu) = 1$, we have seen is $4\pi^4/15$; it must not, however, be assumed that its value is independent of T_1 (for in practice $f(\nu)$ cannot in general be expressed as a function of ν/T_1).

ACKNOWLEDGMENT

I would like to thank Dr. R. A. Smith, Dr. J. G. Daunt and Dr. R. Kompfner for much helpful discussion in the preparation of this paper.

REFERENCES

Born, M., 1944. *Atomic Physics* (London: Blackie and Son).

Daunt, J. G., 1945. *S.R.E. Report*, C. L. Misc., **52**. (Unpublished.)

Daunt, J. G., 1947. (In course of publication).

Einstein, A., 1909. *Phys. Z.*, **10**, 185.

Fowler, R. H., 1936. *Statistical Mechanics* (Cambridge: The University Press).

Fürth, R., 1928. *Z. Phys.*, **50**, 310.

Lorentz, H. A., 1916. *Les Théories Statistiques en Thermodynamique* (Leipzig: Teubner).

Milatz, J. H. W., 1943. Brownsche Beweging. *Proceedings of the XXIXth Netherland Scientific Congress.*

Roberts, J. K., 1928. *Heat and Thermodynamics* (London: Blackie and Son.)

Wilson, H. A., 1944. *Modern Physics* (London: Blackie and Son).

REFLECTING MICROSCOPES

By C. R. BURCH, F.R.S.,
University of Bristol

MS. received 5 *March* 1946 ; *in revised form* 16 *September* 1946

ABSTRACT. The history of reflecting microscopes is reviewed, and the leading Schwarzschild design formulae quoted. Two reflecting objectives are described, and photomicrographs taken with one of these are reproduced.

§1. HISTORICAL INTRODUCTION

REFLECTING microscopes are almost as old as reflecting telescopes, for it was early realized that if the light be sent backwards through a telescope objective, it becomes a microscope objective, albeit of inconveniently great focal length. Newton made a reflecting microscope objective consisting of an ellipsoidal mirror and a diagonal flat; a two-mirror objective more akin to the Cassegrain type is described in Smith's *Complete System of Optics* (1738). These "compound reflecting engiscopes"—as such microscopes were called—were not aplanatic, so that although they could give good images at the low numerical aperture of 0·05 – 0·1 (which for many years satisfied astronomers in their corresponding telescope objectives) they gave seriously comatic off-axis images when the N.A. was raised. Further, a large fraction of the N.A. was necessarily obstructed by the shadow of the second mirror : this could in certain circumstances lead to undesirable effects such as a spurious doubling of the number of lines in the image of a grating. Finally, no satisfactory technique had been developed, either for making or testing the required aspheric surfaces. Accordingly, reflecting micro-objectives passed into oblivion. However, in 1905, Schwarzschild gave the analytical solution of the problem of designing an aplanatic (i.e. spherically corrected and coma-free) two-mirror telescope objective. Algebraically this is of course also the solution of the problem of the two-mirror micro-objective when used at infinite or very great tube length, though the range of numerical values of Schwarzschild's two design parameters m, ϵ, corresponding to practical designs of micro-objective, differs from that corresponding to practical designs of telescope objective. Schwarzschild's classic memoir seems to have attracted relatively little attention, for in 1922 the problem of the two-mirror aplanat was attacked independently by Chrétien, at the instigation of G. W. Ritchey, who had been struck by the fact that the 60″ telescope at Mount Wilson was more nearly aplanatic when used as a Cassegrain than as a Newtonian. "He suspected," writes Chrétien, "that the introduction of the hyperbolic mirror produced a kind of compensation of the aberration of the parabolic mirror : he asked me to study this system theoretically, and in particular, to enquire if it was not possible to improve the imaging properties further by freely forsaking the parabolic and hyperbolic shapes previously given to mirrors."

One of the first to take up afresh the study of reflecting microscopes seems to have been D. D. Maksutov, who, in U.S.S.R. patent No. 40859 (1932), mentions the use of the sphere-cardioid aplanatic pair as a micro-objective, and shows an extremely ingenious "solid" reflecting objective in which the same air-glass surface, after acting as "first mirror" by total internal reflection, lets out the image formed by the second (silvered) surface at normal incidence. His work does not appear to have been suggested by that of Schwarzschild or Chrétien, whose general formulae are not quoted.

More recently he has designed reflecting objectives composed of two spherical mirrors with one or more relatively weak lenses on the image side. Photomicrographs taken in U.V. with one of these objectives have been published by E. M. Brumberg and others. The combination of reflection and refraction has been exploited in a different way by B. K. Johnson, who has published U.V. photomicrographs taken with a reflecting objective analogous to the Mangin mirror, while E. H. Linfoot has combined more reflection with less refraction in two reflecting objectives of Schmidt type, which he showed at the Physical Society's exhibition in 1939.

§ 2. SCHWARZSCHILD APLANATS

Consider the reflecting objective shown in figure 1, consisting of a concave mirror of paraxial radius ρ and a convex of radius R separated by a distance ϵ, the unit of length being the focal length of the combination. Parallel light is imaged at distance m from the pole of the ρ-mirror. ρ, R, ϵ, m are all positive in the diagram. Schwarzschild showed that the shapes of the two mirrors may be so defined that the system is spherically corrected and satisfies the sine condition, and he solved the resulting differential equations in finite form, obtaining for the ρ-mirror the equation

$$\frac{1}{r}=\frac{t}{\epsilon}+\frac{1}{m}\,\frac{\left(1-\dfrac{t}{\epsilon}\right)^{\frac{1}{1-\varepsilon}}}{(1-t)^{\frac{\varepsilon}{1-\varepsilon}}} \qquad \ldots\ldots(1)$$

where

$$t=\sin^2 \tfrac{1}{2}u \qquad \ldots\ldots(2)$$

and r is the length of a ray leaving the object-point at angle u to its incidence-point on the ρ-mirror.

The incidence point of this ray on the second mirror may conveniently be expressed in Cartesian coordinates:

$$\left.\begin{aligned} x&=\epsilon-(1-t)\left[\frac{r(\epsilon-2t)+t}{\epsilon-rt}\right]\ldots\\ y&=2t^{\frac{1}{2}}(1-t)^{\frac{1}{2}}=\sin u\ldots \end{aligned}\right\} \qquad \ldots\ldots(3)$$

The incidence angle i' of this ray on the first (ρ) mirror is given by

$$\tan i'=\frac{(\epsilon-r+1-t)}{\epsilon-t}\left(\frac{t}{1-t}\right)^{\frac{1}{2}} \qquad \ldots\ldots(4)$$

Equations (1) to (4) define the aplanat in terms of m and ϵ, so that we have to choose numerical values for these parameters.

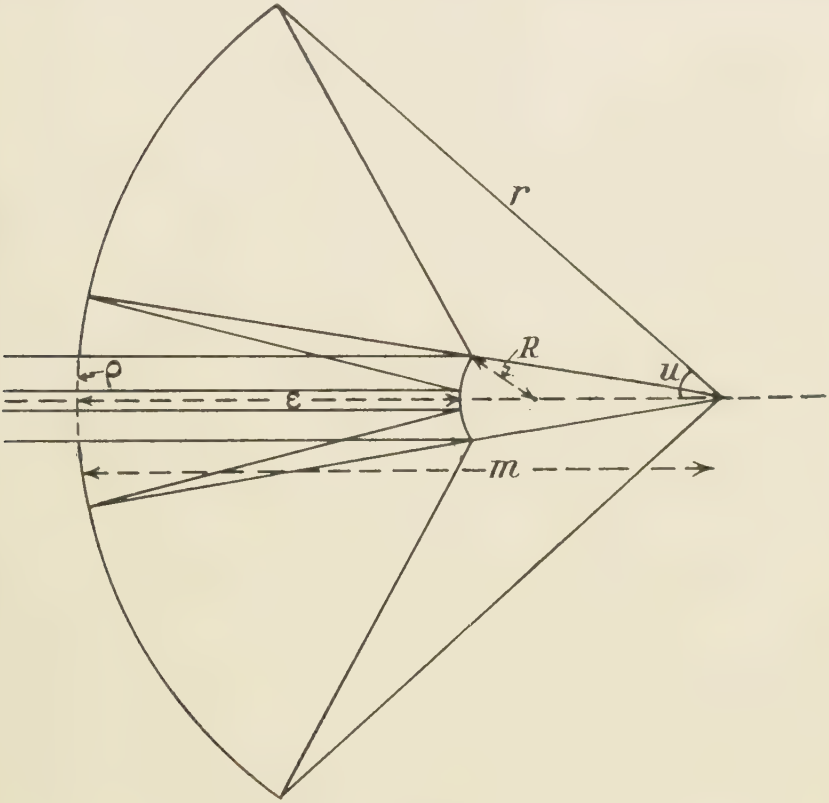


Figure 1. Schwarzschild aplanat.

I do not propose to discuss the design-problem exhaustively, but shall give, for he convenience of those wishing to work on these objectives, the principal approxinate formulae with which the designer will find himself concerned. We may be nfluenced by the coefficient of astigmatism or that of field curvature or by the bstruction ratio or by manufacturing considerations. Schwarzschild showed nat the astigmatic interfocal distance which would be needed in object-space to ive parallel rays at (small) angle θ to the axis in image-space is

$$\text{A.I.D.} = \frac{2-\epsilon}{2m} . \theta^2, \qquad \ldots\ldots(5)$$

nat is, $\frac{2-\epsilon}{2m}$ times the "thin lens" value. He showed also that the surface idway between the focal lines will be curved

$$\frac{(m-1)^2+\epsilon(1-\epsilon)}{2m\epsilon} . \theta^2 \qquad \ldots\ldots(6)$$

vay from the Gaussian image plane.

The fraction of the N.A. shadowed out by the R-mirror is approximately

$$\frac{1}{m-\epsilon} \quad \text{or} \quad m-\epsilon, \qquad \ldots\ldots(7)$$

whichever is less than unity. (In the latter case the light must first pass through a hole in the R-mirror). This approximation contains no unused margin allowance, and is a little over optimistic when the N.A. is not small.

For ease of manufacture we would prefer one mirror to be spherical if this is permissible. Accordingly we seek series approximations to equations (1) and (3) Schwarzschild gives for the ρ-mirror the Cartesian series

$$x = m + \left[\frac{1-m}{\epsilon} - 1\right]\frac{y^2}{4m} - \left\{\frac{(1-m)^2}{\epsilon} - \frac{(1-m)}{\epsilon} + \frac{1}{2\epsilon}\right\}\frac{y^4}{16m^3}$$
$$+ \left[2\left(\frac{1-m}{\epsilon}\right)^3 - 2\left(\frac{1-m}{\epsilon}\right)^2 + \frac{2}{\epsilon}\left(\frac{1-m}{\epsilon}\right) - \frac{1+\epsilon}{6\epsilon^2}\right]\frac{y^6}{64m^5}, \text{ etc.} \quad \ldots\ldots(8)$$

If we disregard terms above y^6, the "non-elliptic" part of this is

$$\mathrm{P}_{\mathrm{NE}}(x) = \frac{-m}{128\epsilon^2}\left[\frac{1+\epsilon}{3} + \frac{\epsilon}{1-m-\epsilon}\right]\left(\frac{y}{m}\right)^6, \quad \ldots\ldots(9)$$

while the remainder is an ellipse of paraxial radius

$$\rho = \frac{2m\epsilon}{m+\epsilon-1} \quad \ldots\ldots(10)$$

and of eccentricity e' given by

$$e'^2 = \left(\frac{1-m+\epsilon}{1-m-\epsilon}\right)^2 + \frac{2\epsilon^2}{(1-m-\epsilon)^3}, \quad = e_0'^2 - \frac{1}{4\epsilon}\left(\frac{\rho}{m}\right)^3, \quad \ldots\ldots(11)$$

where e_0' is the eccentricity of that ellipse whose foci are the object-point and paraxial intermediate image-point.

For the R-mirror Schwarzschild gives $x = A + By^2 + Cy^4 + Dy^6 + Ey^8$ where

$$\left.\begin{array}{l} A = m-\epsilon\ ; \quad B = \dfrac{1-m}{4\epsilon}\ ; \quad C = -\dfrac{1}{8}\cdot\dfrac{m}{4\epsilon}\ ; \\ D = -\dfrac{1}{96}\cdot\dfrac{1+4\epsilon}{\epsilon}\cdot\dfrac{m}{4\epsilon}; \quad E = -\dfrac{1}{1536}\cdot\dfrac{2+11\epsilon+30\epsilon^2}{\epsilon^2}\cdot\dfrac{m}{4\epsilon}. \end{array}\right\} \quad \ldots\ldots(12)$$

This is approximately an ellipse of paraxial radius

$$R = \frac{2\epsilon}{m-1} \quad \ldots\ldots(13)$$

and of eccentricity e given by

$$e^2 = 1 + \frac{2\epsilon^2 m}{(1-m)^3}. \quad \ldots\ldots(14)$$

The non-elliptic part of the sixth power coefficient is

$$\mathrm{P}_{\mathrm{NE}}(D) = D - \frac{2C^2}{B}. \quad \ldots\ldots(15)$$

Let us summarize the main design-restrictions which these formulae imply. For greatest ease of manufacture we equate e' and e simultaneously to zero. This gives $\epsilon = 2$, so that the system is anastigmatic, and $m = 2+\sqrt{5}$, $\rho = 1+\sqrt{5}$; $R = \sqrt{5}-1$, a monocentric objective. The price that we must pay for having zero first coefficient of asphericity on both mirrors—so that they may be spherical up to N.A. approaching 0·5—is the high intrinsic obstruction ratio, $1/\sqrt{5}$. If we wish to retain anastigmatism, we must set $\epsilon=2$, and, except in the case just considered, e' and e both differ from zero for every value of m, so that both mirrors must be asphérized. If we are prepared to give up exact anastigmatism, and be content with aplanatism only, we may set e zero, giving $\epsilon^2=\dfrac{(m-1)^3}{2m}$, and select m so as to give as small an obstruction ratio, $\dfrac{1}{m-\epsilon}$, as we please. This gives an aplanat in which—provided terms involving the sixth power of the N.A. may be neglected—the R-mirror may be spherical. Alternatively we can set e' zero, but this leads to a higher astigmatic coefficient for a given obstruction ratio. I show in a separate paper that when the design is optimized, the R-mirror may be spherical in an objective of small obstruction ratio, up to 0·65 N.A. If we ask—may the R-mirror ever be exactly spherical, up to N.A. unity, the answer is yes, provided $\epsilon=\frac{1}{2}$, $m=2$. The Schwarzschild concave curve then becomes a cardioid, and the combination is the well-known sphere-cardioid pair used in aplanatic dark-ground condensers, usually in approximation as two spheres.

§ 3. EXPERIMENTAL

I have made two "through-type" reflecting objectives, which, though they are corrected for finite tube-lengths, may be regarded as approximating to Schwarzschild aplanats having respectively $\rho=4{\cdot}31$ cm., $R=0{\cdot}85$ cm., $m=8$, $\epsilon=4$, $f=0{\cdot}75$ cm.; N.A. ·58, tube-length 32 cm., magnification ×47, and $\rho=5$ cm., $R=0{\cdot}397$ cm., $m=20$, $\epsilon=13$, $f=0{\cdot}3$ cm.; N.A. ·65, tube-length 30 cm., magnification × 100. To this objective I have added a normal-incidence oil-immersion lens, the surface of which is spherical and concentric with the axial object point. This raises the N.A. to 0·98, and the magnification to about 152, the magnification now being proportional to the refractive index of the lens.

The first objective has both mirrors aspherized: the concave mirror was first aspherized so as to annul spherical aberration for a certain arbitrarily chosen separation between the mirrors and object position, after which the convex mirror was moved axially with respect to the concave so as to annul off-axis coma (the final image position being unchanged). The spherical aberration re-introduced by this change was then annulled by aspherizing the convex mirror. This procedure produces a sufficient approximation to aplanatism: a more elaborate procedure would be needed for N.A.$>$0·7.

Local figuring was needed to restore revolution symmetry lost during aspherizing, and the residual error is of "cobbled pavement" type, of the order of $\frac{1}{8}$ fringe by double transmission. The leading off-axis error for image points 1 cm. off-axis is a fraction of a fringe of astigmatism.

The second objective has a nominally spherical convex mirror, the spacing between the mirrors being adjusted to give the best compromise between high-order and Seidel coma; the leading error for image points $\frac{1}{2}$ cm. off-axis consists of a fraction of a fringe of compromise—comatic retardation of the edge of the wave-front remote from the axis: for image points 1 cm. off-axis the leading error is astigmatism. This objective also has a small fraction of a fringe of "cobble pavement" error.

I find it advisable, with both objectives, to stop out, in the substage condense that part of the N.A. which is obstructed in the objective. If this is not done, th contrast is poor when the substage N.A. is reduced, as is to be expected, for if th substage N.A. is reduced until only the cone obstructed by the objective is supplie the conditions are those of "patch-stop dark-ground illumination" and blac objects are seen bright on a black ground. [If a dark-ground effect is desired, find it better to use a standard dark-ground illuminator supplying a cone lyin outside that used by the objective, as the diffraction rings surrounding images ar then much less prominent.]

The visual performance of both objectives used with full substage aperture o stained specimens seems comparable in contrast and resolving power with that o refracting objectives of equal N.A.

Some idea of the photographic performance may be obtained from figures 2–6 which are reproduced from enlargements made by Mr. H. Busby of films taken by Dr. G. P. Occhialini with the second objective. For their kindly interest an enthusiastic co-operation I record my grateful thanks.

Figures 2 and 3, taken without the oil immersion component, show "thorium stars" ×1140, and treponema pallida, silver-stained by Levaditi's method × 950, 0·65 N.A. ; 0·65 substage N.A.

Figures 4, 5 and 6 were taken with oil immersion : figure 4 shows part of the long chromosome of Drosophila, ×1420, ·98 N.A., ·45 substage N.A. ; figure 5 shows treponema pallida, Levaditi stained, ×2230, and figure 6 gonococci in pus cells, stained with methylene blue, ×2230, 0·98 N.A., 0·98 substage N.A. The light source in all cases was a half-watt projector lamp, diffused by ground glass, with a green filter. No eyepiece was used, the film being placed at the primary image.

REFERENCES

Brumberg, 1939. *C.R. Acad. Sci., U.S.S.R.*, **25**, 473 ; 1941 a. *Ibid.*, **31**, 658 ; 1941 b. *Ibid.*, **32**, 486 ; 1943. *Nature, Lond.*, **152**, 357.

Burch, 1943. *Nature, Lond.*, **152**, 748.

Burch, 1943. *Proc. Phys. Soc.*, **55**, 433. 1945. *Ibid.*, **57**, 567.

Chrétien, 1922. *Rev. d'Optique*, **1** 232.

Clay, 1939. *J. Sci. Instrum.*, **16**, 49.

Johnson, 1941. *Proc. Phys. Soc.*, **53**, 714.

Linfoot, 1938. *J. Sci. Instrum.*, **15**, 405.

Maksutov. U.S.S.R. Patent No. 40859.

Schwarzschild, 1905. *Theorie der Spiegelteleskop* (Göttingen Observatory).

Smith, 1738. *Complete System of Optics*.

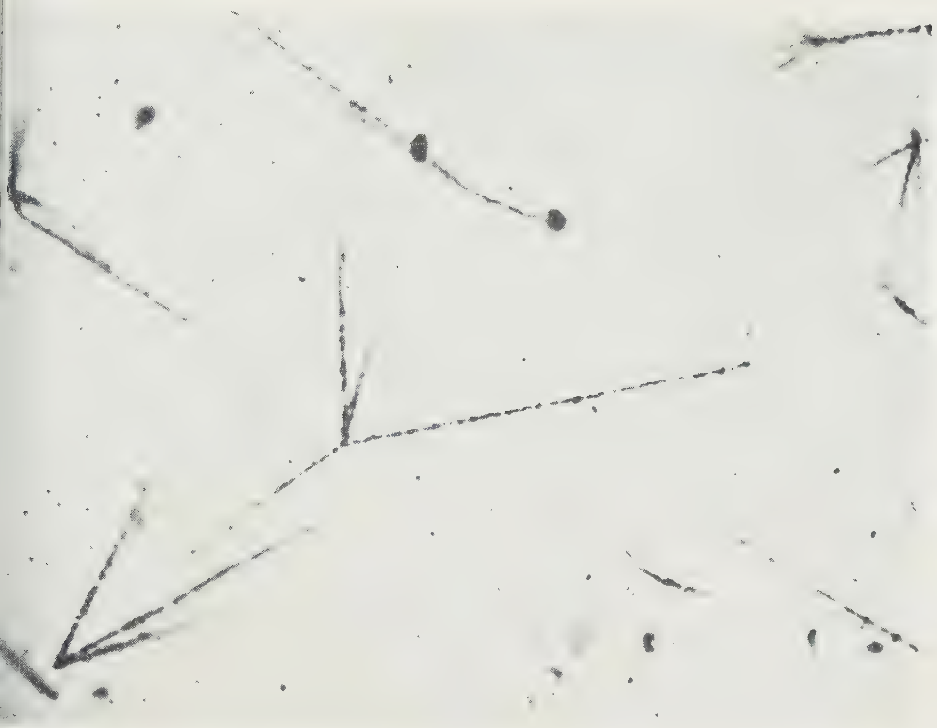


Figure 2.

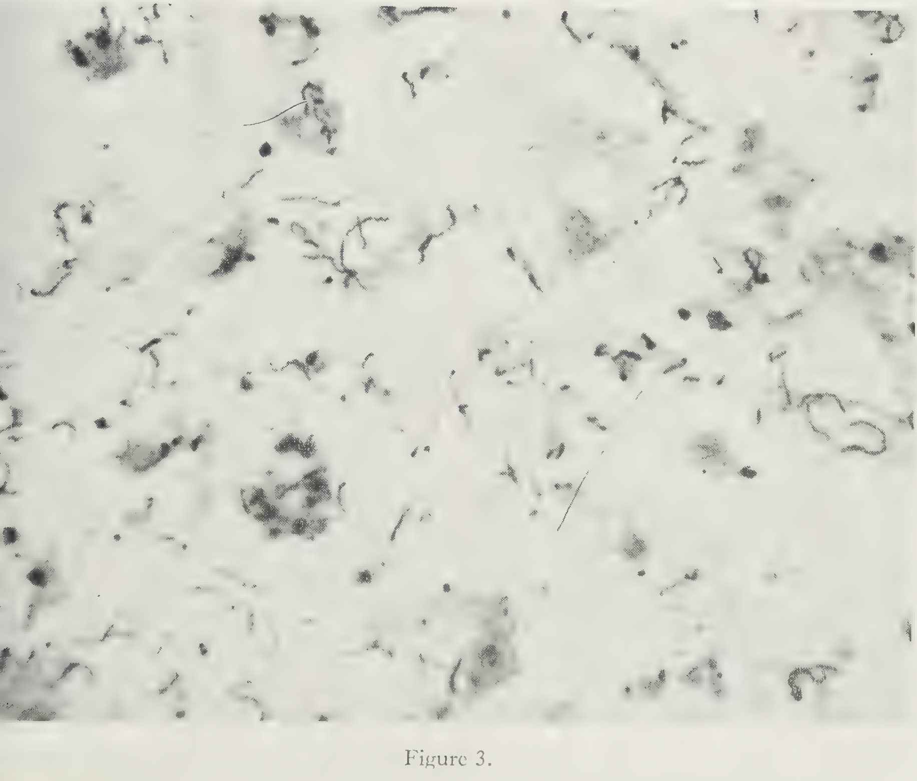


Figure 3.

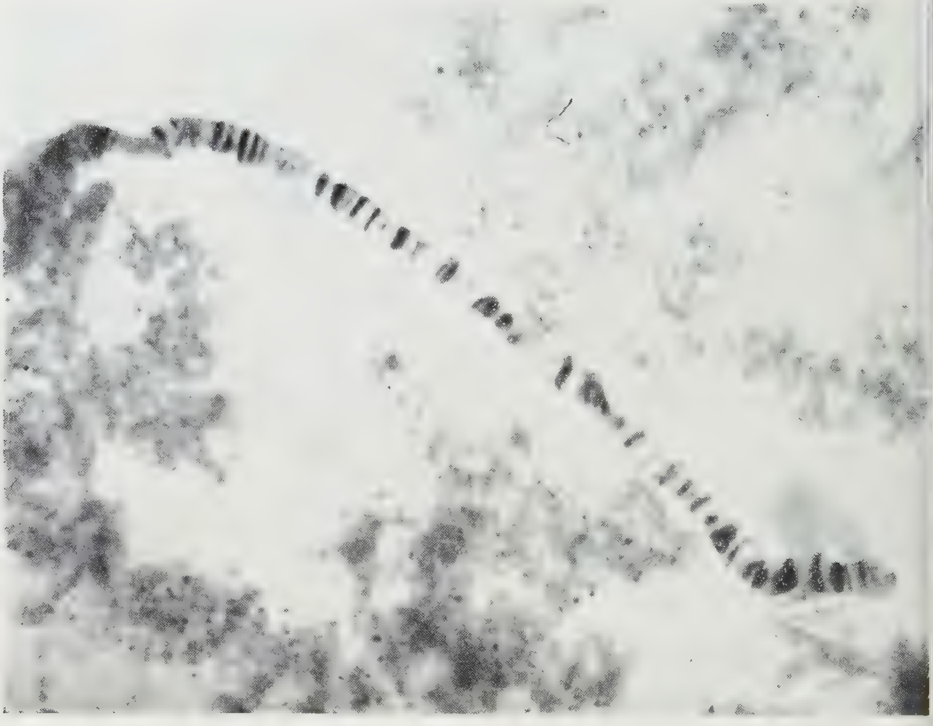


Figure 4.

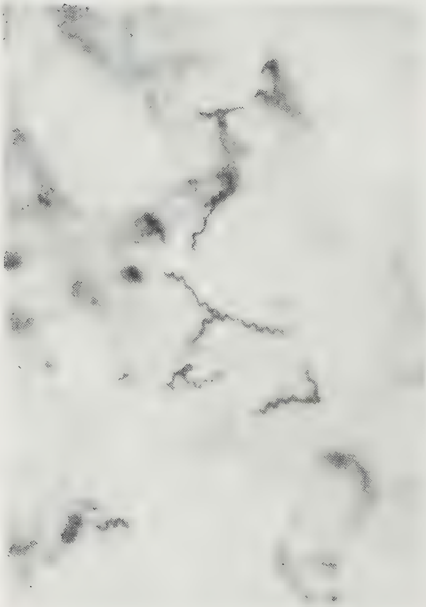


Figure 5.

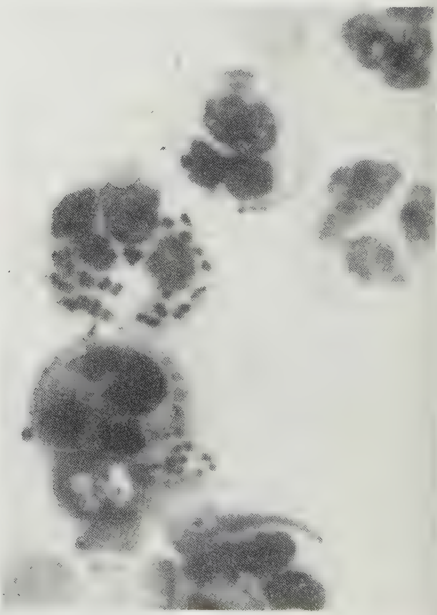


Figure 6.

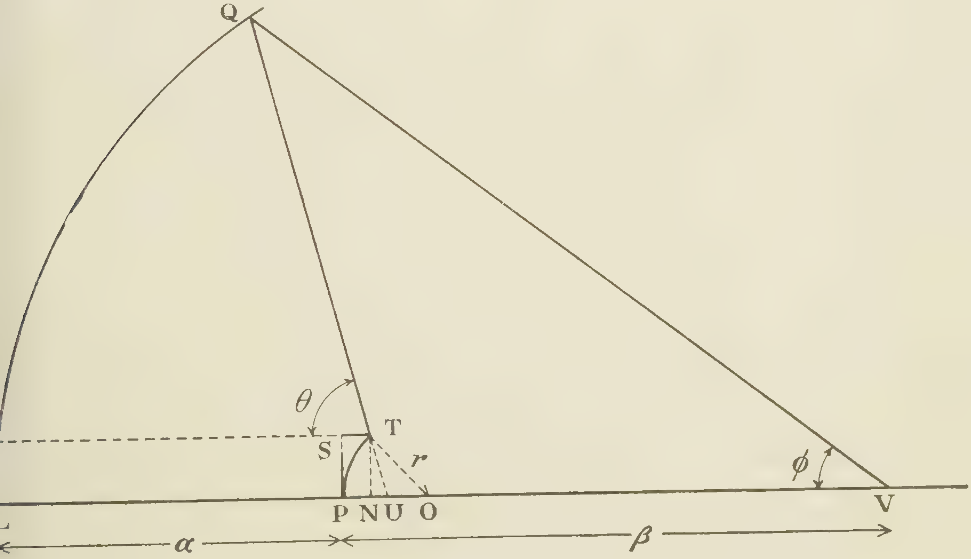
SEMI-APLANAT REFLECTING MICROSCOPES

By C. R. BURCH, F.R.S.,
Bristol University

MS. received 5 *March* 1946 : *in revised form* 9 *August* 1947

BSTRACT. Numerical values are derived for the offence against the sine condition of flecting micro-objectives composed of one aspheric concave mirror and one spherical nvex mirror. Numerical apertures up to 0·65 are found practicable.

THE purpose of this paper is to derive numerical values for the offence against the sine condition ("OSC") of "semi-aplanat" reflecting microscope objectives composed of an aspheric concave "primary" mirror and a pherical convex "secondary", placed so as to form the final image at infinite r very great tube length.



Semi-aplanat reflecting micro-objective.

Such an objective is shown in the figure. We suppose that the concave mirror figured so that the final image is free from spherical aberration. Then the ath length VQTS, from the object-point V to a point S on a plane through e pole of the convex mirror, is independent of the position of the incidence-oint, Q, on the concave mirror. Further, if r be the radius of the convex mirror, the angle of the intermediate ray, $LP=\alpha$, $PV=\beta$, e the eccentricity and the semi-major axis of the ellipse of which Q is a point and U and V foci, we ave (see figure)

$$VQTS=\beta+2\alpha, \qquad \ldots\ldots(1)$$

$$\frac{r}{2}\left(2-\sec\frac{\theta}{2}\right)=\mathrm{PU}, \qquad \ldots\ldots(2)$$

$$r\cos\frac{\theta}{2}=\mathrm{NO}, \qquad \ldots\ldots(3)$$

$$\beta-\frac{r}{2}\left[2-\sec\frac{\theta}{2}\right]=\mathrm{VU}=ae, \qquad \ldots\ldots(4)$$

and

$$\beta+2\alpha-\frac{r}{2}\left[2-2\cos\frac{\theta}{2}-\sec\frac{\theta}{2}\right]=\mathrm{VL}+\mathrm{LP}-\mathrm{PU}+\mathrm{NO},\ =\beta+2\alpha-\mathrm{ST}+\mathrm{TU}$$
$$=\mathrm{VQ}+\mathrm{QU}=2a. \qquad \ldots\ldots(5)$$

Hence

$$e=\frac{\beta-\frac{r}{2}\left[2-\sec\frac{\theta}{2}\right]}{\beta+2\alpha-\frac{r}{2}\left[2-2\cos\frac{\theta}{2}-\sec\frac{\theta}{2}\right]}. \qquad \ldots\ldots(6)$$

From the geometry of the ellipse,

$$\sin\phi=\frac{2(1-e^2)\cos\frac{\theta}{2}\sin\frac{\theta}{2}}{(1+e)^2-4e\sin^2\frac{\theta}{2}}. \qquad \ldots\ldots(7)$$

The height of the parallel ray imaged at angle ϕ is

$$r\sin\frac{\theta}{2}, \qquad \ldots\ldots(8)$$

Set

$$K(\theta)=\frac{(1-e^2)\cos\frac{\theta}{2}}{(1+e)^2-4e\sin^2\frac{\theta}{2}}. \qquad \ldots\ldots(9)$$

Then the offence against the sine condition is

$$\mathrm{OSC}=\frac{\sin\phi_{\mathrm{ideal}}-\sin\phi_{\mathrm{actual}}}{\sin\phi_{\mathrm{ideal}}}=\frac{K(0)-K(\theta)}{K(0)}. \qquad \ldots\ldots(10)$$

Substituting for e from (6) we can expand $K(\theta)$ as a power series in θ^2 with leading terms

$$K(\theta)=\frac{1-e_0}{1+e_0}\left[1+\frac{\theta^2}{8}\left\{\frac{e^2{}_0-6e_0+1}{(1+e_0)^2}+\frac{re_0}{(1-e_0)\left(\beta-\frac{r}{2}\right)}\right\}\right] \ldots\ldots(11)$$

so that if the θ^2 coefficient of $K(\theta)$ is to be zero—i.e. if the objective is to have zero Seidel coefficient of coma—we must have

$$\frac{\beta}{r}=\frac{1}{2}+\frac{e_0(1+e_0)^2}{(1-e_0)(-e_0{}^2+6e_0-1)}. \qquad \ldots\ldots(12)$$

We would like to have $\beta \gg r$, so as to have a negligible fraction of the N.A. obstructed by the shadow of the convex mirror. Consider therefore first the limiting case $\beta/r=\infty$, which implies $e_\theta=e_0, =(1+e_0{}^2)/6, =3-2\sqrt{2}=\cdot 1718\ldots$ The magnification of the intermediate image, $(1-e_0)/(1+e_0)$, then becomes $1/\sqrt{2}$.

$K(\theta)$ now becomes

$$\frac{\cos\frac{\theta}{2}}{\sqrt{2}\left(1-\frac{1}{2}\sin^2\frac{\theta}{2}\right)}=\frac{1}{\sqrt{2}}\left(1-\frac{\theta^4}{128}+\text{etc.}\right) \qquad \ldots\ldots (13)$$

so that if we set the OSC tolerance at 1/400 and neglect higher terms, we must imit θ to roughly $1/\sqrt{2}$ radian, and the NA ($\sin\phi$) to about 1/2. We can reduce the maximum OSC by a re-choice of e_0, provided we do not demand that the θ^2-coefficient of OSC be zero. When $\beta/r=\infty$, we have for any e_0

$$K(\theta)=\frac{1-e_0}{1+e_0}\cdot\frac{\cos\frac{\theta}{2}}{1-\frac{4e_0}{(1+e_0)^2}\sin^2\frac{\theta}{2}} \qquad \ldots\ldots (14)$$

so that we should choose e_0 so as to make $1-\frac{4e_0}{(1+e)^2}\sin^2\frac{\theta}{2}$ imitate $\cos\frac{\theta}{2}$ as closely as possible. Let us make the imitation exact at $\theta=60°$; then we must set $e_0=\cdot1896\ldots$, and the imitation is found to be correct to about 1·1 parts in 400 up to $\theta=60°$, for which $\sin\phi=\cdot6812$, which is also the value of $(1-e_0)/(1+e_0)$, the magnification of the intermediate image. This suggested that a useful approximation to optimal design for N.A. 0·65 or so might be obtained when β/r is finite by a corresponding increase in e_0—say, a 10% increase—over the value given by (12). Three values of e_0 were accordingly chosen in this way for the three values 5·75, 3·50 and 2·75 of β/r and the OSC was computed for a number of values of θ by means of (6), (9), and (10). The results, together with the obstructed fraction of the N.A., i.e. $\frac{(1+e_0)}{2(1-e_0)}\cdot\frac{r}{\beta}$ and the associated values of α/r, are given in table 1.

Table 1

θ	$\sin\phi$	OSC	$\sin\phi$	OSC	$\sin\phi$	OSC	$\sin\phi$	OSC
10	0·1188	$0{\cdot}27\times10^{-3}$	0·1163	$0{\cdot}23\times10^{-3}$	0·1138	$0{\cdot}22\times10^{-3}$	0·1114	$0{\cdot}21\times10^{-3}$
20	0·2368	0·99	0·2317	0·88	0·2269	0·82	0·2222	0·81
30	0·3533	1·90	0·3457	1·68	0·3385	1·61	0·3314	1·63
40	0·4617	2·55	0·4571	2·40	0·4476	2·29	0·4383	2·39
50	0·5771	2·26	0·5647	2·19	0·5531	2·33	0·5417	2·70
55	0·6300	1·46	0·6166	1·62	0·5904	1·92	0 5917	2·50
60	0·6812	0 0	0·6670	0·44	0·6540	0·66	0·6405	1·94
65	0·7304	−2·21	0·7154	−1·37	0·7014	0·34	0·6875	0·91
70	0·7771	−5·47	0·7617	−4·04	0·7471	−2·71	0·7328	−0·68
e_0	0·1896		0·20		0·21		0·2201	
β/r	inf		5·75		3·50		2·75	
α/r	inf		10·0		5·143		3·486	
ruction	0·0		$\frac{1}{7{\cdot}66}$		$\frac{1}{4{\cdot}57}$		$\frac{1}{3{\cdot}5}$	

These objectives should therefore be reasonably coma-free up to 0·65 N.A. t is fortunate that the sign of the OSC is such that it can be corrected by a uitable turn-*down* of the edge of the convex mirror, together with a corresponding eduction of asphericity of the concave.

THE HEATING OR COOLING OF A SOLID SPHERE IN A WELL-STIRRED FLUID

By S. PATERSON,
Imperial Chemical Industries Ltd., Explosives Division

MS. received 26 *August* 1946

ABSTRACT. A solid sphere of uniform initial temperature is heated or cooled in finite mass of fluid, well stirred and externally insulated. Temperature continuity assumed at the surface. Solutions are presented suitable for all values of time, radiu and conductivity, and of the ratio of heat capacities of sphere and fluid. Numeric results are given, and certain mathematical relations noted.

---

§ 1. SERIES SOLUTION

A SOLID sphere of radius a, diffusivity κ, heat capacity c_1 and temperature zer is plunged at time $t=0$ into a mass of fluid of heat capacity c_2 at temperatur T_0. The surface temperature of the sphere is assumed equal at all subsequent times to that of the fluid, which is so well stirred or of such high conductivit that its temperature is always uniform throughout. The outer boundary of th fluid is impervious to heat.

Then the temperature T at a distance R from the centre of the sphere determined by

$$\left.\begin{array}{l} \dfrac{\partial}{\partial t}(RT)=\kappa\dfrac{\partial^2}{\partial R^2}(RT), \quad R<a, \quad t>0,\\ \qquad T=0, \quad R<a, \quad t=0,\\ \qquad T=T', \quad R=a, \quad t>0,\\ \qquad T'=T_0, \quad t=0,\\ \dfrac{3\kappa c_1}{a}\dfrac{\partial T}{\partial R}=-c_2\dfrac{dT'}{dt}, \quad R=a, \quad t>0, \end{array}\right\} \quad \ldots\ldots(1)$$

where $T'(t)$ is the temperature of the fluid.

For convenience, let $R/a\equiv r$, $\kappa t/a^2\equiv\tau$, $RT/aT_0\equiv u$, $T'/T_0\equiv u'$; also $c_1/c_2\equiv w$ Then the equations are

$$\left.\begin{array}{ll} \dfrac{\partial u}{\partial\tau}=\dfrac{\partial^2 u}{\partial r^2}, \quad r<1, \quad \tau>0, & u=u', \quad r=1, \quad \tau>0,\\ & u'=1, \quad \tau=0,\\ u=0, \quad r<1, \quad \tau=0, & u=0, \quad r=0, \quad \text{all } \tau,\\ \multicolumn{2}{c}{3w\left(\dfrac{\partial u}{\partial r}-\dfrac{u}{r}\right)+\dfrac{du'}{d\tau}=0, \quad r=1, \quad \tau>0.} \end{array}\right\} \quad \ldots\ldots(2)$$

If $p\equiv -s^2$ is the Heaviside operator $\partial/\partial\tau$, these equations transform into

$$\left.\begin{array}{ll} \dfrac{\partial^2 u}{\partial r^2}=-s^2 u, & u=0, \quad r=0,\\ 3w\left(\dfrac{\partial u}{\partial r}-\dfrac{u}{r}\right)-s^2(u-1)=0, & r=1. \end{array}\right\} \quad \ldots\ldots(3)$$

The solution is readily found to be

$$u=\frac{s^2\sin rs}{3w(\sin s-s\cos s)+s^2\sin s} \qquad \ldots\ldots(4)$$

$$\equiv f(s)/F(s),\ \text{say}.$$

The poles s_n of u are the roots of

$$s\cot s=1+s^2/3w, \qquad \ldots\ldots(5)$$

nd apart from a simple pole at zero consist of real and distinct pairs, positive and negative. The solution (4) therefore transforms into

$$u=\underset{s\to 0}{\mathcal{L}}\frac{f(s)}{F(s)}+2\sum_{n=1}^{\infty}\frac{f(s_n)}{s_n F'(s_n)}e^{-s_n^2\tau}.$$

Hence
$$u=\frac{r}{w+1}+2\sum_{1}^{\infty}\frac{\sin rs_n}{\sin s_n}\cdot\frac{e^{-s_n^2\tau}}{3(w+1)+s_n^2/3w}, \qquad \ldots\ldots(6)$$

where s_n is the nth positive non-zero root of (5). The corresponding fluid temperature is

$$u'=(u)_{r=1}=\frac{1}{w+1}+2\sum_{1}^{\infty}\frac{e^{-s_n^2\tau}}{3(w+1)+s_n^2/3w}. \qquad \ldots\ldots(7)$$

The above formal solution can be justified by verifying that (6) and (7) do in fact satisfy (2).

When $\tau=\infty$, $u=u'r=r/(w+1)$, as is required. The fraction F of the final heat transfer which has taken place by time t is therefore

$$F=\frac{1-u'}{1-1/(w+1)}, \qquad \ldots\ldots(8)$$

that is,

$$F=1-\frac{2}{3w}\sum_{1}^{\infty}\frac{e^{-s_n^2\tau}}{1+s_n^2/9w(1+w)}, \qquad \ldots\ldots(9)$$

(5), (6) and (7) correspond, for the present condition of uniform initial temperature, to the general series solution given by Peddie (1901). When the fluid has infinite heat capacity, that is, when the surface temperature of the sphere is maintained at T_0, they reduce to

$$u=r+\frac{2}{\pi}\sum_{1}^{\infty}(-1)^n\frac{\sin n\pi r}{n}e^{-n^2\pi^2\tau}, \qquad \ldots\ldots(10)$$

$$u'=1,$$

which is a familiar case. (9) then becomes

$$F=1-\frac{6}{\pi^2}\sum_{1}^{\infty}\frac{e^{-n^2\pi^2\tau}}{n^2}. \qquad \ldots\ldots(11)$$

The nth root s_n of (5) can be seen from a graph of the functions $\tan s$ and $s(1+s^2/3w)$ to lie between $n\pi$ and $(n+\frac{1}{2})\pi$. For $n<(\sqrt{3w}/\pi)-\frac{1}{2}$, s_n approaches $(n+\frac{1}{2})\pi$ as n is increased, thereafter returning towards $n\pi$; thus, if w is small, s_n approaches $n\pi$ from the outset. s_n can be calculated without difficulty by successive approximation. If n/w is sufficiently large, the following expansion is also useful:

$$s_n = n\pi + \frac{3w}{n\pi} - \frac{9w^2(w+2)}{(n\pi)^3} + \dots \qquad \dots\dots(12)$$

The first three roots, which are more than we shall require, are presented in table 1 for a wide range of w.

Table 1. The first three positive roots of $s \cot s = 1 + s^2/3\omega$

w	s_1	s_2	s_3	w	s_1	s_2	s_3
∞	4·712	7·854	10·996	5·0	4·236	7·296	10·329
1000	4·492	7·722	10·899	2·0	3·972	6·938	9·940
500	4·490	7·720	10·895	1·0	3·726	6·681	9·714
400	4·487	7·719	10·893	0·5	3·506	6·502	9·576
300	4·486	7·716	10·890	0·2	3·312	6·376	9·487
200	4·485	7·713	10·885	0·1	3·233	6·330	9·454
100	4·479	7·702	10·866	0·05	3·188	6·307	9·439
50	4·464	7·674	10·830	0·02	3·161	6·293	9·430
20	4·421	7·601	10·759	0·01	3·152	6·290	9·426
10	4·352	7·490	10·573	0·0	3·142	6·283	9·425

The series in (6), (7), etc. no doubt always converge. The rapidity of convergence, however, varies markedly with τ. Thus in (7), if $\tau \geqslant 1$, the exponent $-s_n^2\tau$ increases numerically by a factor of at least $(2n+1)\pi^2$, and the entire series is negligible, so that $u = r/(w+1)$, $u' = 1/(w+1)$ and $F = 1$, irrespective of w; but when $\tau < 1$, the speed of convergence falls off very quickly, at least two terms being required for $\tau = 0·1$, eleven for $\tau = 0·01$ and over one hundred for $\tau = 0·001$. Since for certain purposes it may be necessary to consider values of τ as low as, say, 10^{-10}, the series solution evidently cannot be regarded in practice as covering more than a fraction of the range.

§ 2. ALTERNATIVE SOLUTIONS

We return to (4), substitute $-q^2$ for s^2, and expand in powers of e^{-q}. This gives

$$u = \frac{q^2}{q^2 + 3wq - 3w}\{e^{-(1-r)q} - e^{-(1+r)q} + Qe^{-(3-r)q} - Qe^{-(3+r)q} + \dots\} \qquad \dots\dots(13)$$

and

$$u' = \frac{q^2}{q^2 + 3wq - 3w}\{1 + (Q-1)e^{-2q} + Q(Q-1)e^{-4q} + \dots\}, \qquad \dots\dots(14)$$

where

$$Q \equiv \frac{q^2 - 3wq - 3w}{q^2 + 3wq - 3w}. \qquad \dots\dots(15)$$

Since e^{-2q} transforms into $1 - \operatorname{erf} 1/\sqrt{\tau}$, we may expect that the bracketed terms beyond the first two in (13) and the first in (14) will be negligible when $1 - \operatorname{erf} 1/\sqrt{\tau}$ is small, say $< 0·1\%$ for $\tau \leqslant 0·1$. Then (13) and (14) should be suitable for calculation precisely in the range in which (6) and (7) are unsuitable.

In order to obtain an approximation for sufficiently small τ, we may expand $q^2/(q^2+3wq-3w)$ in descending powers of q, and interpret operators of the type $q^{-n}e^{-Aq}$ in terms of Hartree's repeated error function integrals (Hartree, 1936). Thus, if $P \equiv 3w$,

$$\frac{q^2 e^{-(1\pm r)q}}{q^2+Pq-P} = \Phi_0\left(\frac{1\pm r}{2\sqrt{\tau}}\right) - P(2\sqrt{\tau})\,\Phi_1\left(\frac{1\pm r}{2\sqrt{\tau}}\right) + (P+P^2)(2\sqrt{\tau})^2\,\Phi_2\left(\frac{1\pm r}{2\sqrt{\tau}}\right) - \dots, \qquad \dots\dots(16)$$

where

$$\Phi_n(x) \equiv \int_x^\infty \Phi_{n-1}(\xi)\,d\xi$$

and

$$\Phi_0(x) \equiv 1 - \operatorname{erf} x.$$

When $r=1$, the terms in $1+r$ are relatively small, and the first approximation to u' is

$$u' = 1 - P\,.\,2\sqrt{\tau/\pi} + (P+P^2)\tau - \dots \qquad \dots\dots(17)$$

For sufficiently small $w\sqrt{\tau}$, say $<10^{-2}$, these expansions are useful, particularly (16), when $r \ll 1$, but they do not cover the required range. We shall therefore derive alternative expressions.

The terms in $e^{-(1\pm r)q}$ from (13) transform into

$$\mp\frac{1}{\pi i}\int_M \frac{e^{\mu^2\tau-(1\pm r)\mu}\mu\,d\mu}{(\mu-\alpha)(\mu-\beta)},$$

respectively, where α, β are the roots of $\mu^2+3w\mu-3w=0$, and M is a path in the μ-plane from $\infty e^{-i\pi/4}$ to $\infty e^{i\pi/4}$ passing a finite distance to the right of $|\alpha|$ and $|\beta|$. The terms therefore yield

$$u \simeq \frac{1}{\alpha-\beta}\{\alpha[E_{1-r}(\alpha)-E_{1+r}(\alpha)] - \beta[E_{1-r}(\beta)-E_{1+r}(\beta)]\}, \qquad \dots\dots(18)$$

where

$$\left.\begin{aligned} E_y(x) &\equiv \frac{1}{\pi i}\int_{M'} \frac{e^{\mu^2\tau-y\mu}\,d\mu}{\mu-x} \\ &= e^{x^2\tau-xy}\left[1-\operatorname{erf}\left(\frac{y}{2\sqrt{\tau}}-x\sqrt{\tau}\right)\right] \\ &= e^{-y^2/4\tau}G\left(\frac{y}{2\sqrt{\tau}}-x\sqrt{\tau}\right), \end{aligned}\right\} \qquad \dots\dots(19)$$

where

$$G(z) \equiv e^{z^2}(1-\operatorname{erf} z).$$

(18) and (19) give a first approximation to u. It can be further simplified for particular ranges of r or w. Thus, if r is small, (18) becomes

$$u \simeq \frac{2r}{\alpha-\beta}\left\{\alpha^2E_1(\alpha) - \beta^2E_1(\beta) + \frac{\alpha-\beta}{\sqrt{\pi\tau}}e^{-1/4\pi\tau}\right\}. \qquad \dots\dots(20)$$

The relative temperature u/r at the centre is obtained at once from (20 Again, if r is nearly 1, the first bracketed term in (13) will, for small τ, be larg compared with the second, so that

$$u \simeq \frac{1}{\alpha-\beta}\{\alpha E_{1-r}(\alpha) - \beta E_{1-r}(\beta)\}. \qquad \ldots\ldots(21)$$

A first approximation to the fluid temperature is obtained from the first term of (13), or by setting $r=1$ in (21). Thus

$$u' \simeq \frac{1}{\alpha-\beta}\{\alpha E_0(\alpha) - \beta E_0(\beta)\}. \qquad \ldots\ldots(22)$$

Finally, when $w=0$,

$$\begin{aligned} u &\simeq E_{1-r}(0) - E_{1+r}(0) \\ &= \operatorname{erf}\frac{1+r}{2\sqrt{\tau}} - \operatorname{erf}\frac{1-r}{2\sqrt{\tau}}, \end{aligned} \qquad \ldots\ldots(23)$$

$$\begin{aligned} F &= \frac{3(q\cosh q - \sinh q)}{q^2 \sinh q} \\ &= -3\tau + 6\sqrt{\tau/\pi} + 12\sqrt{\tau}\sum_1^{\infty}\overline{\Phi}_1\left(\frac{n}{\sqrt{\tau}}\right). \end{aligned} \qquad \ldots\ldots(24)$$

(23) and (24) correspond to (10) and (11). (24) converges very rapidly indeed for $\tau \leqslant 0{\cdot}1$, giving $F = 6\sqrt{\tau/\pi} - 3\tau$ to better than 1 in 30,000, while for $\tau \geqslant 0{\cdot}1$ two terms at most are required of the series in (11).

No matter how small $1-r$ is, short of being absolutely zero, the expressions given by (18), (21) and (23) tend to zero with τ, as required. On the other hand, (22), which is of course continuous with (21) for $\tau > 0$, tends to unity as τ approaches zero. The solutions thus correctly reproduce the required continuity for $\tau > 0$ and discontinuity at $\tau = 0$. It is not the least advantage of the operational method, as remarked by Jaeger (1945), that this is made possible in cases like the present without any special analytical device.

It is of interest to confirm that the second term in (14) is in fact negligible. This term transforms into

$$-\frac{6w}{\pi i}\int_M \frac{e^{\mu^2\tau - 2\mu}\mu^2\, d\mu}{(\mu-\alpha)^2(\mu-\beta)^2}.$$

We may resolve the integrand into partial fractions, and integrate the terms in $(\mu-\alpha)^{-2}$ and $(\mu-\beta)^{-2}$ by parts. There results

$$\begin{aligned} + \frac{12w}{(\alpha-\beta)^3}\Big\{&[\alpha\beta - \alpha^2(\alpha-\beta)(\alpha\tau-1)]E_2(\alpha) - [\alpha\beta + \beta^2(\alpha-\beta)(\beta\tau-1)]E_2(\beta) \\ &- (\alpha-\beta)(\alpha^2+\beta^2)\sqrt{\frac{\tau}{\pi}}\,e^{-\frac{1}{\tau}}\Big\}. \end{aligned} \qquad \ldots\ldots(25)$$

Let $\alpha \equiv \tfrac{1}{2}(-3w + \sqrt{9w^2+12w})$ and $\beta \equiv \tfrac{1}{2}(-3w - \sqrt{9w^2+12w})$. Clearly $0 < \alpha < 1$ and $\beta < 0$. Values of α and β for a wide range of w are shown in table 2. If, then, $\tau < 0{\cdot}1$, $1/\sqrt{\tau} - \alpha\sqrt{\tau}$ and $1/\sqrt{\tau} - \beta\sqrt{\tau}$ are of the order of 3, or greater. Hence, using the asymptotic series for erf, we have

$$\left.\begin{aligned} E_2(\alpha) &\sim e^{-1/\tau}/\sqrt{\pi}(1/\sqrt{\tau} - \alpha\sqrt{\tau}), \\ E_2(\beta) &\sim e^{-1/\tau}/\sqrt{\pi}(1/\sqrt{\tau} - \beta\sqrt{\tau}). \end{aligned}\right\} \qquad \ldots\ldots(26)$$

Table 2. The roots α, β of $x^2+3wx-3w=0$

w	α	β	w	α	β
1000	0·9997	−3001	5·00	0·9410	−15·94
500	0·9993	−1501	2·00	0·8730	− 6·873
400	0·9992	−1201	1·00	0·7913	− 3·791
300	0·9989	− 901·0	0·50	0·6861	− 2·186
200	0·9983	− 601·0	0·20	0·5307	− 1·131
100	0·9967	− 301·0	0·10	0·4179	− 0·7179
50	0·9934	− 151·0	0·05	0·3195	− 0·4695
20	0·9839	− 60·98	0·02	0·2168	− 0·2768
10	0·9693	− 30·97	0·01	0·1583	− 0·1883

The factor $e^{-1/\tau}$ already guarantees that the term in $E_2(\alpha)$ in (23) is negligible ; nd the same can at once be shown to be true of the remaining two terms in virtue f the cancellation of the potentially large β^3.

Thus, for $\tau \leqslant 0{\cdot}1$, u' is given to a close approximation by (22), that is by

$$u' \simeq \frac{1}{\alpha-\beta}\{\alpha e^{\alpha^2\tau}(1+\operatorname{erf}\alpha\sqrt{\tau})-\beta e^{\beta^2\tau}(1+\operatorname{erf}\beta\sqrt{\tau})\}, \qquad \ldots\ldots(27)$$

nd F follows from (8).

By an argument precisely similar to the above, it can be shown that for $\tau \leqslant 0{\cdot}1$ he terms in (13) beyond the first two are also negligible. So also, of course, is he second itself if r is near to 1.

§ 3. NUMERICAL RESULTS

u' and F have been computed for $w=1000$ to $w=0$ and $\tau=10^{-12}$ to $\tau=1$. 27) was used for $\tau \leqslant 0{\cdot}1$, and (7) for $\tau \geqslant 0{\cdot}1$. A comparison between the two types of solution is thus provided at $\tau=0{\cdot}1$. It will be seen from table 3 that the agreement is close, considering that only four-figure tables were used. The course of F is shown in figure 1. Here an additional check was made at $w=0$, $\tau=0{\cdot}1$ and ·01, (11) yielding $F=0{\cdot}7705$, $0{\cdot}3084$ and (24) $F=0{\cdot}7705$, $0{\cdot}3085$. Finally, s an illustration, the temperature field was calculated for $w=1$, $\tau=10^{-4}$, 10^{-2}, 0^{-1} by (18) and for $\tau=0{\cdot}1$, 1 by (6). Very close agreement was found at $\tau=0{\cdot}1$. The distribution is shown in figure 2. Except at the centre u/r rises temporarily bove its final value.

§ 4. APPENDIX

Several interesting relations emerge from the above. Thus, from (9), since $F \to 0$ as $\tau \to 0$,

$$\sum_{n=1}^{\infty}\frac{1}{s_n^2+9w(1+w)}=\frac{1}{6(1+w)}, \qquad \ldots\ldots(28)$$

$_n$ being the nth positive non-zero root of (5). $\Sigma\, 1/n^2=\pi^2/6$ is of course a special ase, given by (11).

Again, from (11) and (24),

$$\sum_{1}^{\infty}\frac{e^{-n^2\pi^2\tau}}{n^2}=\frac{\pi^2}{6}\left\{1+3\tau-6\sqrt{\frac{\tau}{\pi}}-12\sqrt{\tau}\sum_{1}^{\infty}\Phi_1\left(\frac{n}{\sqrt{\tau}}\right)\right\}. \qquad \ldots\ldots(29)$$

Table 3. $u' \equiv T'/T_0$ for various values of $w \equiv c_1/c_2$ and $\tau \equiv \kappa t/a^2$

	Calculated from Equation (27)												Calculated from Equation (7)	
τ / w	10^{-12}	10^{-11}	10^{-10}	10^{-9}	10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{-1}	$\geqslant 1$
1000·00	0·997	0·989	0·967	0·901	0·735	0·442	0·178	0·060	0·019	0·006	0·002	0·001	0·001	0·000
500·00		0·995	0·983	0·949	0·851	0·629	0·322	0·117	0·038	0·012	0·005	0·002	0·002	0·002
400·00		0·996	0·987	0·959	0·878	0·683	0·370	0·145	0·048	0·016	0·006	0·003	0·003	0·002
300·00			0·990	0·969	0·906	0·746	0·457	0·188	0·063	0·021	0·008	0·004	0·004	0·003
200·00			0·993	0·979	0·936	0·818	0·568	0·272	0·093	0·031	0·011	0·005	0·005	0·005
100·00			0·997	0·988	0·967	0·901	0·735	0·443	0·182	0·062	0·022	0·011	0·011	0·010
50·00				0·995	0·984	0·949	0·852	0·630	0·325	0·122	0·044	0·021	0·021	0·020
20·00					0·993	0·979	0·936	0·818	0·571	0·276	0·108	0·052	0·052	0·048
10·00					0·996	0·990	0·967	0·902	0·737	0·451	0·203	0·100	0·100	0·091
5·00						0·995	0·983	0·949	0·852	0·638	0·351	0·185	0·185	0·167
2·00							0·994	0·980	0·937	0·822	0·596	0·369	0·370	0·333
1·00							0·997	0·989	0·967	0·904	0·755	0·548	0·548	0·500
0·50								0·995	0·983	0·950	0·864	0·714	0·714	0·667
0·20									0·994	0·980	0·941	0·865	0·864	0·833
0·10									0·997	0·990	0·970	0·928	0·928	0·909
0·05										0·995	0·985	0·963	0·963	0·952
0·02											0·994	0·985	0·985	0·980
0·01											0·997	0·992	0·992	0·990

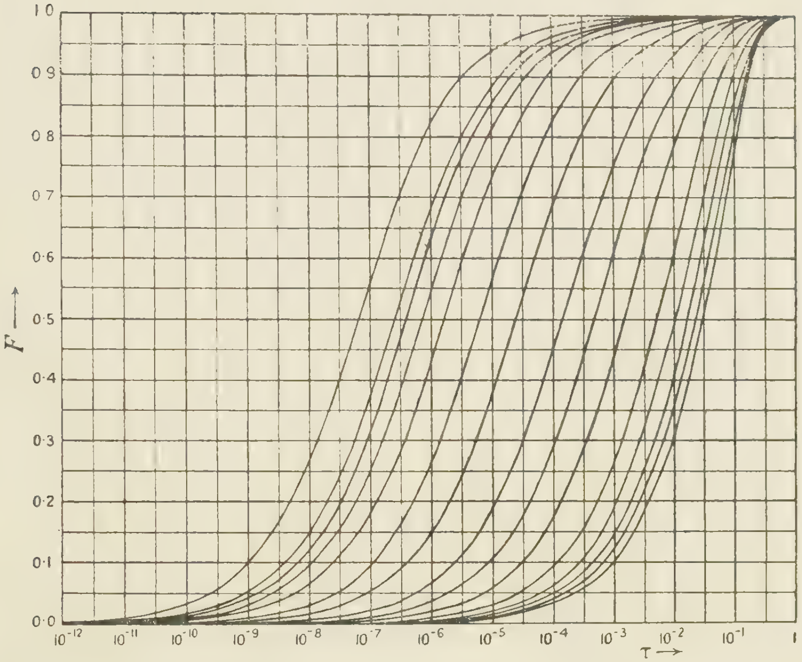


Figure 1. Variation of F with τ for various values of w. From left to right the curves correspond to $w=1000$, 500, 400, 300, 200, 100, 50, 20, 10, 5, 2, 1, 0·5, 0·2, 0.

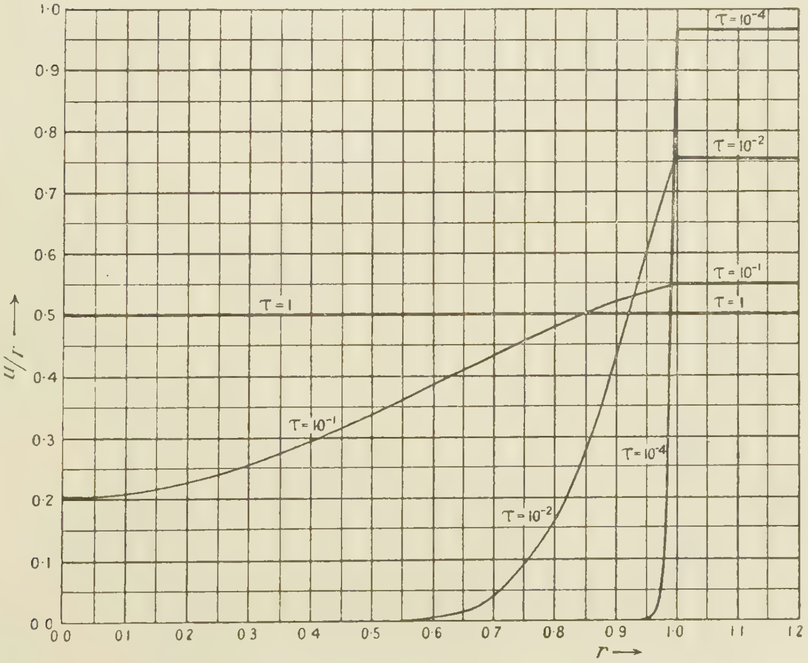


Figure 2. Variation of u/r with r for $w=1$ and various values of τ.

This yields, very closely, for $0<x\leqslant 1$,

$$\sum_1^\infty \frac{e^{\;n^2x}}{n^2} \simeq \frac{\pi^2}{6} - \sqrt{\pi x} + \frac{x}{2} \qquad \ldots\ldots(30)$$

and

$$\sum_1^\infty \Phi_1(nx) \simeq \frac{1}{4x} - \frac{1}{2\sqrt{\pi}} + \frac{x}{12}. \qquad \ldots\ldots(31)$$

By differentiating (29) and (30) with regard to x, or integrating from x to ∞ a series of useful formulae can be derived. For example,

$$\sum_1^\infty e^{-n^2x} = \tfrac{1}{2}\left(\sqrt{\frac{\pi}{x}} - 1\right) + \sqrt{\frac{\pi}{x}} \sum_1^\infty e^{-n^2\pi^2/x} \qquad \ldots\ldots(32)$$

and

$$\sum_1^\infty \frac{e^{-n^2x}}{n^4} = \frac{\pi^4}{90} - \frac{\pi^2 x}{6} + \frac{2\sqrt{\pi}}{3}x^{3/2} - \frac{x^2}{4} + \pi(2\sqrt{x})^3 \sum_1^\infty \Phi_3\left(\frac{n\pi}{\sqrt{x}}\right), \qquad \ldots\ldots(33)$$

the series terms on the right of (32) and (33) being negligible for $0<x\leqslant 1$. (32) is a particular case of "Poisson's Identity".

REFERENCES

HARTREE, D. R., 1936. *Mem. Manch. Lit. Phil. Soc.*, **80,** 85.

JAEGER, J. C., 1945. *Proc. Camb. Phil. Soc.*, **41,** 43.

PEDDIE, W., 1901. *Proc. Edin. Math. Soc.*, **19,** 34.

THE APPLICATION OF IONOSPHERIC DATA TO RADIO COMMUNICATION PROBLEMS: PART II

BY SIR EDWARD APPLETON, F.R.S. AND W. J. G. BEYNON*

* Radio Division, National Physical Laboratory; now at
University College, Swansea

MS. received 30 *July* 1946 .† *Read* 15 *November* 1946

ABSTRACT. Graphs are given from which may be estimated the maximum usable frequency of radio waves reflected by an ionospheric layer in oblique incidence transmission. The curves based on the theory given in Part I of the paper are drawn for such ranges of layer thickness and layer height as are met with in practice. The limitations in the accuracy and applicability of the theory in practice are briefly discussed. Attention is also drawn to the occurrence of abnormal transmission conditions under which long-distance communication via the ionosphere is possible on frequencies exceeding the normally predicted values.

§ 1. INTRODUCTION

In Part I of this paper (Appleton and Beynon, 1940) we described a method of estimating the frequency range of radio waves which are deviated back to the ground when incident obliquely on the ionosphere, and which are, therefore, suitable for communication over various distances. For this purpose formulae ere derived giving directly the value of the maximum usable frequency which is eflected by a thick ionized layer surrounding a curved earth, the distribution of lectronic density in the layer being taken as "parabolic". In this continuation f the paper we exhibit the somewhat complicated formulae previously obtained or the maximum usable frequency in a graphical form, embodying a range of the elevant parameters which our practical experience suggests as adequate. Also, ince the solution previously obtained involves certain approximations, we upplement its graphical expression by a discussion of the accuracy which can be ttained in typical practical cases. Attention is also drawn to the manner in which he analysis can be expressed in the form of transmission curves suitable for the irect estimation of the maximum usable frequency from the vertical-incidence elation between equivalent height of reflection and frequency. In the final ection of the paper we discuss the practical significance of abnormal *E*-layer eflections in effecting long-distance transmission on frequencies exceeding the ormally predicted values.

---

† Sections 1 to 7 of this paper were originally prepared in June 1942 and circulated at that time as a confidential paper by the Radio Research Board.—AUTHORS.

§2. MAXIMUM USABLE FREQUENCY FACTORS

In the theoretical treatment of Part I of this paper the characteristics of the eviating layer in the ionosphere were specified in terms of three quantities, viz.:

(*a*) f^0, the ordinary ray critical frequency of the ionized layer for waves incident normally on it;

(*b*) h_0, the height above ground of the lower edge of the layer; and

(*c*) y_m, the vertical semi-thickness of the layer.

hese quantities can be determined from the curve relating equivalent height h' nd frequency f obtained by the usual method of vertical-incidence radio sounding. he ordinary ray critical frequency of the layer, f^0, can be read off this curve by nspection, while the parameters h_0 and y_m can be determined from the same urve by the method described in the Appendix to Part I of this paper.

Now one of the advantages of the solution previously given is that, by means f it, the relation between the oblique incidence critical frequency $f_{\max}$ (i.e. the aximum usable frequency) and the vertical-incidence critical frequency f^0 can e expressed as

$$f_{\max}/f^0 = x(h_0, y_m, D), \qquad \ldots\ldots(1)$$

here D is the distance of transmission, and where the function x represents what e may call the "maximum usable frequency factor" (M.U.F. factor) by which the xperimentally determined quantity f^0 must be multiplied to give the required uantity $f_{\max}$. The special convenience of the use of (1) in practice arises from he fact that the variation of the function x throughout the hours of the day and he seasons of the year can be ascertained from the results of a preceding year (at ny rate for quiet ionospheric conditions), whereas the fairly substantial changes f f^0 from day to day (and even from hour to hour) are not at present predictable.

The relation between the M.U.F. factor and the distance of transmission D or appropriate ranges of the parameters y_m and h_0 may be exhibited in two series f curves, samples of which are reproduced in small scale * in figures 1 and 2.

---

* A complete set of larger scale graphs will be supplied to *bona fide* users on application to the uperintendent, Radio Division, National Physical Laboratory, Teddington, Middlesex. The opies supplied, unlike the reproductions in this paper, have a graticule.

In figure 1, two sample sets of curves show the relation between the M.U.F factor and D for different appropriate values of the height of maximum laye ionization (y_m+h_0), the ratio y_m/h_0 being constant and equal to 0 and 0·2 respectively. Probably a more convenient representation is that of figure 2, in which th value of the M.U.F. factor is shown as a function of y_m/h_0 for different representativ distances of transmission D and different values of (y_m+h_0).

The wide range of information embodied in figures 1 and 2 should be notec The complete set of curves covers a range of layer heights and thicknesses appropriate to the various ionospheric layers. The limiting curves for layers of zer thickness are identical, as they should be, with those drawn for sharp boundar reflection.

§3 SOME NOTES ON THE PRACTICAL DETERMINATION OF y_m AND h_0

In the Appendix to Part I, a graphical method of determining y_m and h_0 from (h', f) vertical incidence data was described. Briefly, this method consists in th graphical determination of the parameters y_m and h_0 in the equation representin the relation between equivalent height h' and frequency f, namely,

$$h' = h_0 + \frac{y_m}{2} \log_e \frac{f^0+f}{f^0-f}. \qquad \ldots\ldots(2)$$

Now practical use of this method has led us to note two points of importance in it application. In the first place it often happens that the experimental (h', f) curv cannot be represented by (2) over the whole frequency range. In that case it i necessary to choose the constants of the equation so that it represents closely tha part of the (h', f) curve which is most important for our purpose. Secondly, sinc the theory as so far described relates only to the case of a single deviating layer it is necessary to consider the modifications necessary to allow for the influence o refraction in an ionized layer situated below that in which the waves are ultimately reflected. These two matters are considered in greater detail in sections (a) and (b) immediately below.

(a) *The importance of the high-frequency portion of the vertical incidence (h', f) curv*

For a plane ionosphere, the equivalent path of waves of frequency f, inciden at an angle i_0, can readily be expressed in terms of the equivalent path of a frequency $f \cos i_0$ incident normally (Martyn, 1935). Now in Part I of this paper it was shown that a curved ionosphere behaves like a plane ionosphere if we assume that the vertical-incidence critical frequency is reduced to

$$\left(f^0 - \frac{f^2}{f^0} \cdot \frac{y_m}{R+h_0} \sin^2 i_0\right).$$

The equivalent vertical-incidence frequency in the curved ionosphere case, expressed as a fraction of the critical frequency, is therefore given approximately by

$$x_1 = \frac{f \cos i_0}{f^0\left(1 - \frac{f^2}{f^{02}} \frac{y_m}{R+h_0} \sin^2 i_0\right)}. \qquad \ldots\ldots(3)$$

Actual calculations of maximum usable frequencies, using appropriate values of y_m and h_0, show that we are generally concerned with values of x_1 greater than $0{\cdot}9$. Hence it is of great importance to determine y_m and h_0 from that part of the vertical incidence (h', f) curve which indicates approach to penetration (i.e. the parabolic formula should express the distribution of ionization with height near the level of maximum, (y_m+h_0)). For this purpose the equivalent heights should be known accurately for frequencies between $0{\cdot}9f^0$ and f^0.

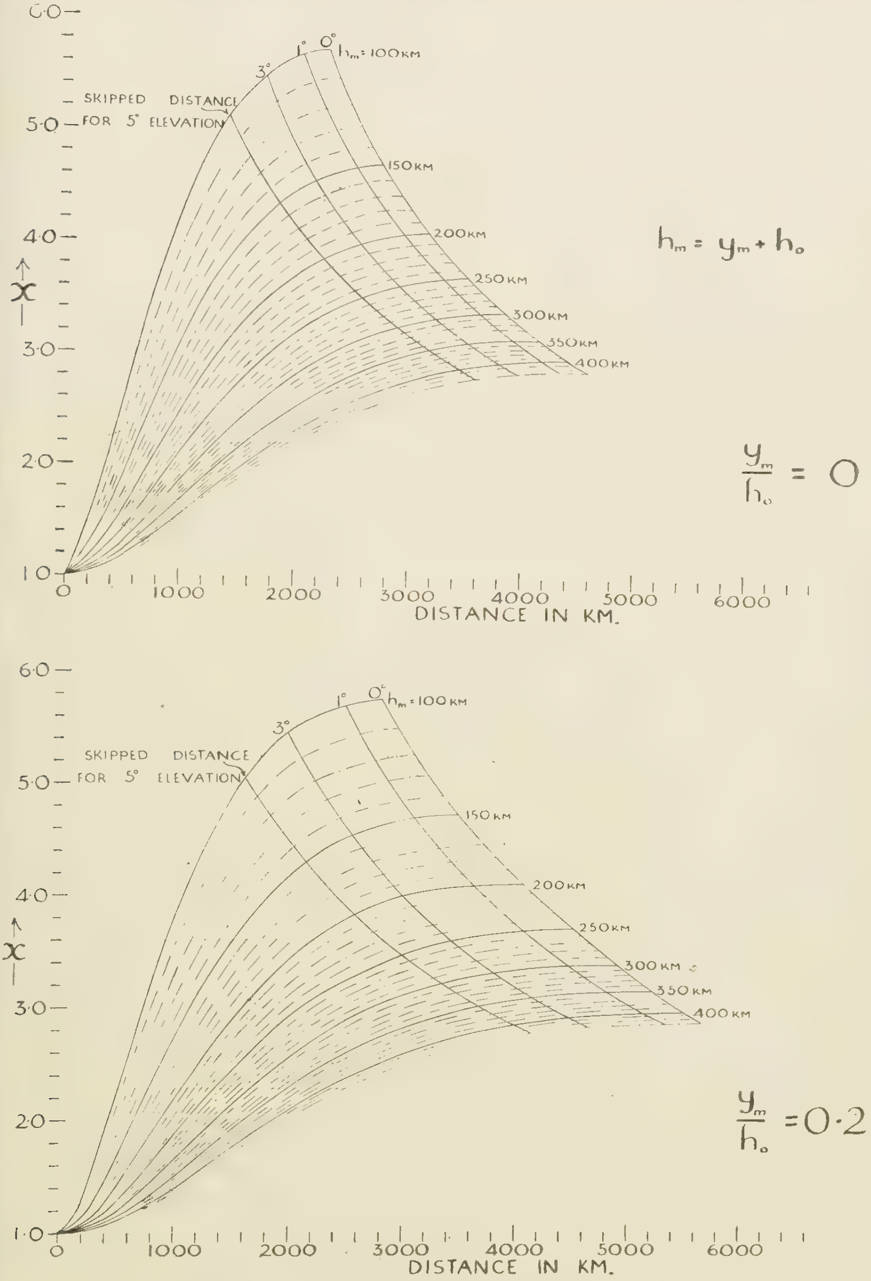


Figure 1. M.U.F. factor curves for a parabolic layer.

(b) *The influence of lower layers on the calculation of the* M.U.F. *factor*

Let us consider the effect of a refracting E-layer below the reflecting F_2 regio the earth being taken as flat. (The argument which follows can be equally we

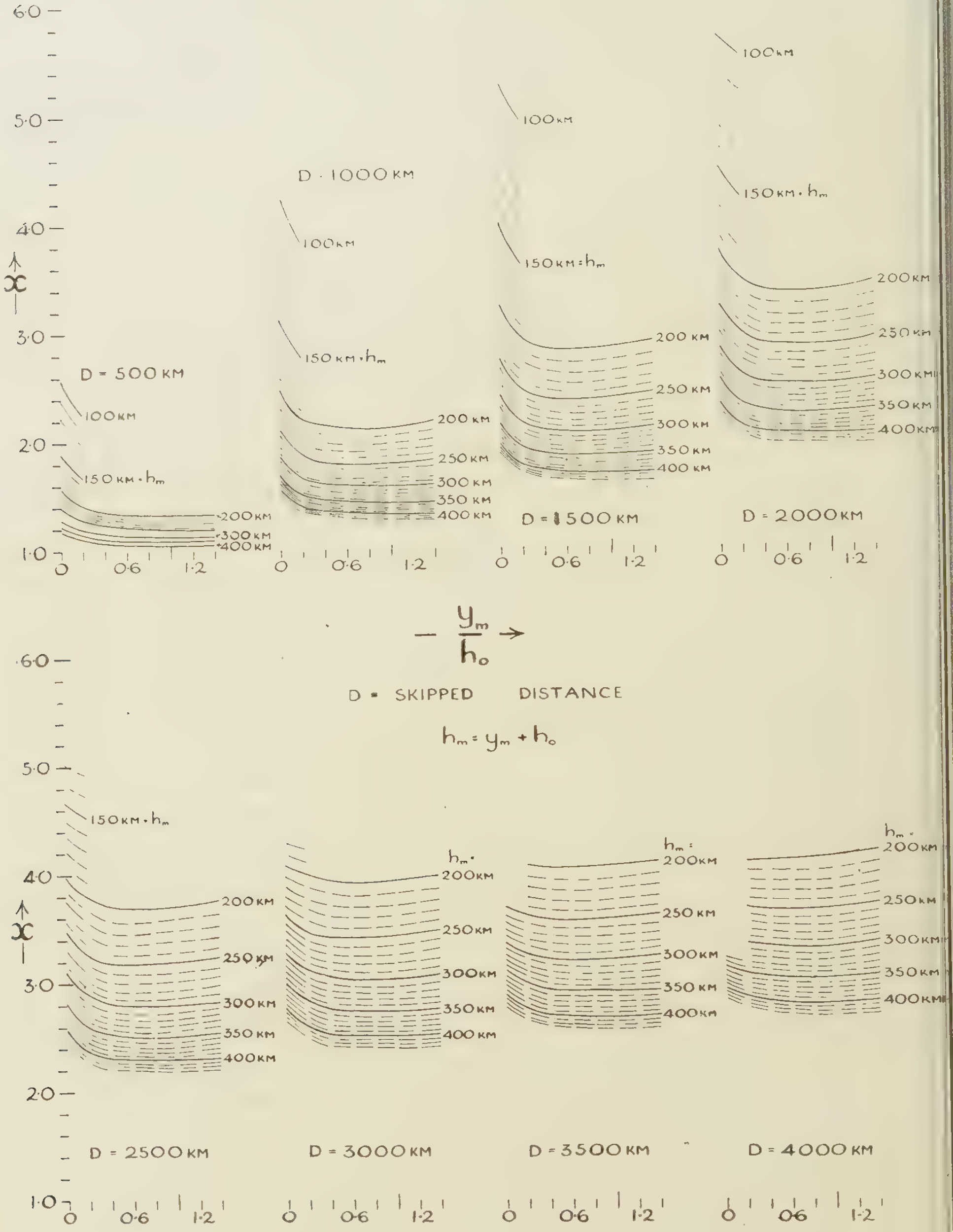


Figure 2. Cross-sections of M.U.F. factor curves at various distances.

applied to the case of region F_1 below region F_2.) We assume that the electroni density distribution with height is of the form shown by the full curves in figure 3 both regions E and F_2 being of parabolic form. Let h, y and f^0, with appropriat suffixes, be the usual constants of the two regions. We consider the trajectory o aves of frequency f incident at an angle i_0 on the E-layer (figure 4). After passing hrough that layer they impinge on the F-layer at the same angle of incidence.

In the absence of the lower layer, the horizontal range would be given by

$$S'R' = y_F x_F \sin i_0 \log_e \frac{1+x_F \cos i_0}{1-x_F \cos i_0} + 2h_F \tan i_0, \qquad \ldots\ldots(4)$$

here $$x_F = f/f_F^0.$$

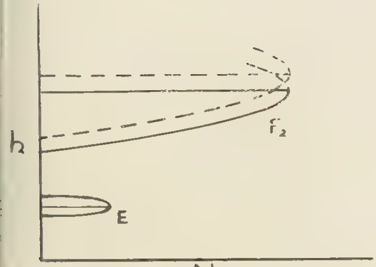


Figure 3.

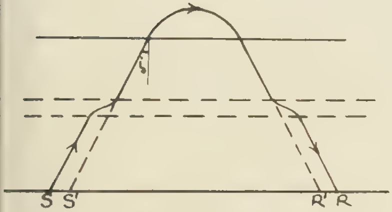


Figure 4.

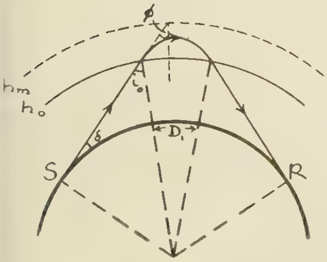


Figure 7.

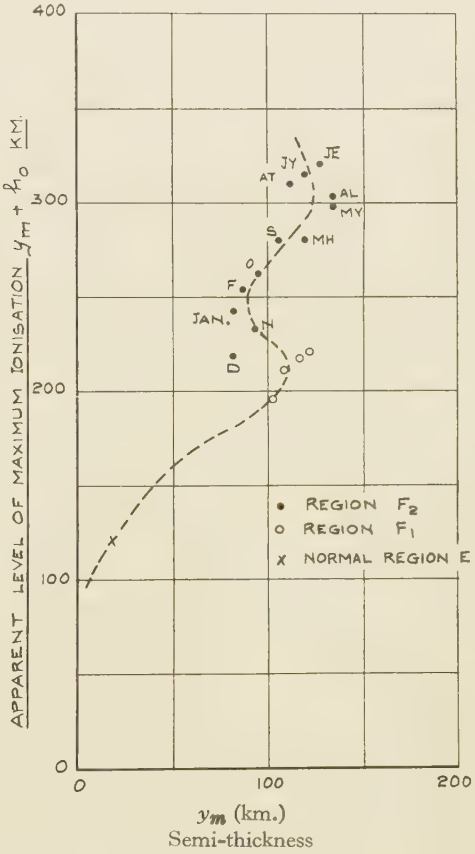


Figure 5. Mean variation of apparent semi-thickness y_m with apparent level of maximum ionization $y_m + h_0$. Monthly mean noon values based on Slough normal incidence data (1943).

With the group-retarding effect of the lower layer included, the actual range is

$$SR = x_F y_F \sin i_0 \log_e \frac{1+x_F \cos i_0}{1-x_F \cos i_0} + 2h_F \tan i_0$$

$$+ 2 \tan i_0 y_E \left\{ x_E \cos i_0 \log_e \frac{x_E \cos i_0 + 1}{x_E \cos i_0 - 1} - 1 \right\}, \qquad \ldots\ldots(5)$$

where $$x_E = f/f_E^0.$$

Now $$x_E \cos i_0 \log_e \frac{x_E \cos i_0 + 1}{x_E \cos i_0 - 1} > 1,$$

so that, for a given frequency incident at a given angle, the effect of a regic beneath the main deviating region is to increase all distances, including "skip distance; that is to say, the M.U.F. of the upper region for a given horizontal ran will be decreased.

Consider now the M.U.F. calculation for these conditions. At vertical inciden the equivalent height for a frequency f is given by

$$h' = \frac{y_F}{2} x_F \log_e \frac{1+x_F}{1-x_F} + h_F - 2y_E + y_E x_E \log_e \frac{x_E+1}{x_E-1}. \qquad \ldots\ldots(6)$$

Expanding the second logarithmic term in (6), we have, since $f/f_E^0 = x_E >$

$$\log_e \frac{f+f_E^0}{f-f_E^0} = \frac{2f_E^0}{f} + \frac{2}{3}\left(\frac{f_E^0}{f}\right)^3 + \ldots.$$

Hence $$h' \simeq \frac{y_F}{2} \cdot \frac{f}{f_F^0} \log_e \frac{f_F^0+f}{f_F^0-f} + h_F + \frac{2}{3} y_E \left(\frac{f_E^0}{f}\right)^2. \qquad \ldots\ldots(7)$$

Now in using the vertical incidence (h', f) curve to deduce the semi-thickne and height of lower boundary of the reflecting layer, we confine our measuremen to frequencies greater than $0{\cdot}9 f_F^0$. For such frequencies we can write $f \simeq f_F^0$, an the effect of the lower layer will be to add a term of approximate magnituc $\frac{2}{3} y_E (f_F^0, f_E^0)^2$ to the calculated value of the height of lower boundary (see dotted curv in figure 3). We have examined theoretically the extent to which the use of th fictitious height in a M.U.F. calculation, for both flat and curved earth case compensates for the effect of the lower layer,* and find that, for values of the laye constants characteristic of normal ionospheric conditions, the errors are rarel likely to exceed 4%.

§4. THE PRACTICAL USE OF THE GRAPHS

As has been explained above, the M.U.F. factor, for any distance of transmissio can be determined when the quantities y_m and h_0 are known. To get the mos reliable estimate of the factor at any time, it is obviously desirable that y_m and h should be determined from vertical-incidence measurements at a place midwa between the sending and receiving stations. Such a course is, however, rarel practicable, and it is usually necessary to forecast the probable values of y_m and h from past experience. As a result of a study made at the Radio Researcl Station, Slough (Lat. 51½° N.), the diurnal, seasonal and sunspot cycle variation of these quantities at this latitude are now known. An account of the result of this investigation will be given in a later communication.

As specimens of the range of values found, we may quote some results fo region F_2 for the year 1944, a year near sunspot minimum. These are given i table 1 below.

It will be seen that the height of the level of maximum ionization $(y_m + h_0$ usually reaches a minimum near noon, and that this level undergoes a seasona change, being higher in summer than in winter.

---

* A note on the formulae, etc. involved in this examination is given in the Appendix.

Table 1

1944	G.M.T.	0000	0600	1200	1800
July	$y_m + h_0$ (km.) y_m/h_0	313 0·43	299 0·54	276 0·57	296 0·58
September	$y_m + h_0$ (km.) y_m/h_0 . .	334 0·43	260 0·56	263 0·58	286 0·54
December	$y_m + h_0$ (km.) y_m/h_0	335 0·54	297 0·52	241 0·47	274 0·59

A further sample of values is shown in figure 5, where the monthly mean noon alues of y_m for the whole of the year 1943 are plotted against the corresponding alues of $y_m + h_0$ obtained from Slough data. Most of the plotted points refer to gion F_2, but some values for region E and region F_1 are also included. The dividual estimates of y_m and $(y_m + h_0)$ often show marked variability, but the ean points shown on the diagram indicate clearly the general tendency for rger values of y_m to be associated with larger values of $(y_m + h_0)$. This general ndency is characteristic of measurements of y_m and h_0 made at all times of day d night, but is particularly well defined in the noon values. The physical gnificance of this increase of thickness of the ionized layer with increasing eight was discussed by one of us some years ago (Appleton, 1939).

Calculations of y_m and h_0 from samples of experimental normal-incidence ', f) curves observed at several widely separated ionospheric observatories dicate that the mean variation of y_m with $(y_m + h_0)$ shown in figure 5 is most robably indicative of that which occurs over a wide range of latitude and longitude.

§ 5. ESTIMATION OF M.U.F. DIRECTLY FROM THE NORMAL INCIDENCE (h',f) CURVE

When the two parameters y_m and h_0 are known, the M.U.F. factor, and hence the .U.F. itself, can be read directly from the set of standard curves. On the other and, if a standard type of (h', f) record is always available, then it is more conenient to be able to read the M.U.F. directly by applying some other set of curves the experimental (h', f) curve. Curves of this type, known as "transmission urves", were first described by Smith (1937) and by Millington (1938), and have rmed the basis of M.U.F. calculations at a number of ionospheric observatories. n the case of the "parabolic layer" method, it is an extremely simple matter to nvert the M.U.F. factor curves described above to a form suitable for direct pplication to any experimental normal incidence (h', f) curves. The transmission urves corresponding to any specified distance of transmission consist in the nvelopes of a series of theoretical normal incidence (h', f) curves drawn for arabolic layers having the constants y_m and h_0 related by the mean curve shown in gure 5. Sets of transmission curves constructed in this way are shown in figure 6.

We have already mentioned that individual estimates of y_m and $(y_m + h_0)$ often how considerable divergence from the mean curve given in figure 5, and it might e expected that the assumption of this fixed relation between these two quantities would result in appreciable errors in the estimates of the M.U.F. Two factor however, tend to make such errors of quite small magnitude. In the first plac as we have already shown, the M.U.F. factor is very largely determined by the magn tude of $(y_m + h_0)$ and depends only to a smaller degree on the actual magnitude the semi-thickness y_m. An examination of the standard M.U.F. curves in conjun tion with our knowledge of the actual variations of y_m, has shown that, as far M.U.F. calculations are concerned, the effect of fluctuations about the values give by the mean curve shown in figure 5 will be small. Secondly, the transmissio curve technique is largely self-compensatory with respect to variations in thickne of the layer, the result being that even the magnitudes of possible errors due assuming a fixed law of variation of thickness with height are, in practice, furth reduced.

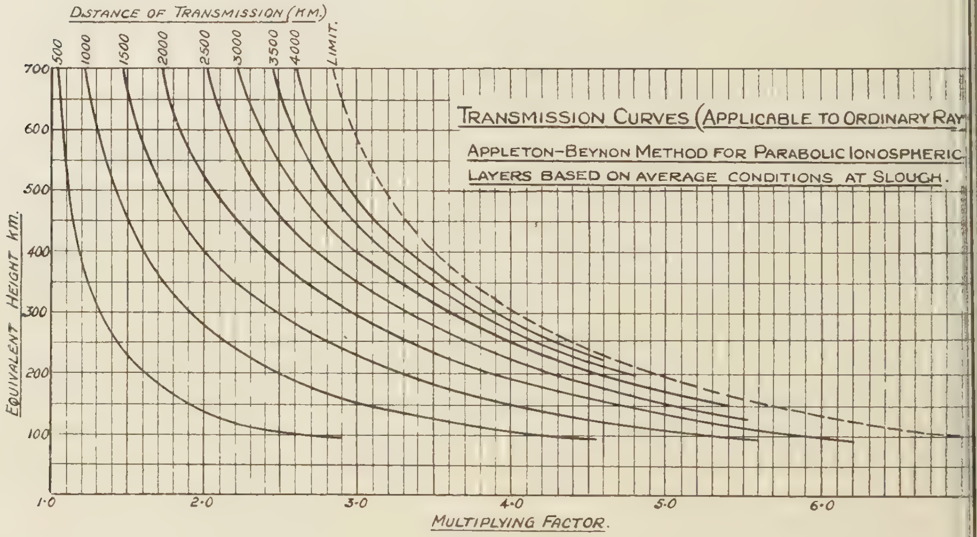


Figure 6.

§6. LIMITATIONS IN THE APPLICABILITY OF THE THEORY

It should be emphasized that our analysis relates strictly to the case of a unifor ionosphere, since we have assumed no variation of ionospheric characteristic over the "part-range" D_1 (figure 7, p. 63). This part-range corresponds to tha portion of the transmission path which lies within the ionosphere itself. Fro the typical numerical calculation given in Part I it will be seen that, for a tot range of 3500 km., D_1 may amount to as much as 1300 km. While it is probabl correct, under most conditions, to consider the values of f^0, h_0 and y_m at the mic point of this part-range as representative of the whole of the ionospheric track we should expect that the results would be less reliable when sunset or sunris conditions obtain over this section of the ionosphere. Under such non-unifor conditions lateral deviation from the great circle path may also occur for certai directions of transmission; in addition, asymmetrical ray tracks are to be expected

Still greater difficulties are encountered when it is attempted to apply th relatively simple theory illustrated above to a case of long-distance transmission i which it is necessary to assume multiple hops. In such cases the changing iono spheric conditions will result in the distance between successive earth reflections being unequal. In such circumstances the application of the above analysis becomes a complicated matter.

§7. THE MAXIMUM FACTOR AT A GIVEN ANGLE OF INCIDENCE

For certain problems, particularly for extreme distance transmission, it may be useful to know the maximum radio-frequency which can be returned from the onosphere at any specified angle of incidence, irrespective of the horizontal distance involved. Referring to figure 7, the part-range D_1, for a frequency f ncident at an angle i_0, is given by

$$D_1 = \frac{R}{R+h_0} \sin i_0 \,.\, x \,.\, y_m \log_e \left\{ \frac{1 - \dfrac{x^2 y_m}{R+h_0} \sin^2 i_0 + x\cos i_0}{1 - \dfrac{x^2 y_m}{R+h_0} \sin^2 i_0 - x\cos i_0} \right\}. \qquad \ldots\ldots(8)$$

For finite values of D_1, it is clear that x must satisfy the inequality

$$\frac{x\cos i_0}{1 - x^2 \dfrac{y_m}{R+h_0} \sin^2 i_0} \leqslant 1. \qquad \ldots\ldots(9)$$

To terms in $\dfrac{1}{R+h_0}$ equation (9) can be written in the form

$$x^2 \left[1 - \sin^2 i_0 \left(\frac{R+h_0}{R+h_0+y_m} \right)^2 \right] \leqslant 1 \qquad \ldots\ldots(10)$$

or

$$x \leqslant \sec\phi, \qquad \ldots\ldots(11)$$

where ϕ is the angle shown in figure 7. Equation (9) can also be reduced to the condition

$$x \leqslant \left\{ \frac{\sqrt{1 + \dfrac{4y_m}{R+h_0} \tan^2 i_0} - 1}{\dfrac{2y_m}{R+h_0} \tan^2 i_0} \right\} \sec i_0. \qquad \ldots\ldots(12)$$

This inequality determines the absolute maximum value of x for any given angle of incidence. It will be noted that for any fixed angle of incidence this absolute maximum factor is determined by the value of the ratio $y_m/(R+h_0)$. Since $h_0 \ll R$, an approximate form of (12) is

$$x \leqslant \left\{ \frac{\sqrt{1 + \dfrac{4y_m}{R} \tan^2 i_0} - 1}{\dfrac{2y_m}{R} \tan^2 i_0} \right\} \sec i_0. \qquad \ldots\ldots(13)$$

Thus, for any given angle of incidence, the maximum value of the multiplying factor x depends principally on y_m, the semi-thickness of the reflecting region. It may be noted that when $R \to \infty$, equation (12) reduces, as it should, to the usual expression for a plane ionosphere, namely $x \leqslant \sec i_0$.

It is often, however, more convenient to express this limiting factor in terms of he angle of emission δ relative to the ground at the sender (see figure 7).

Now

$$\tan^2 i_0 = \frac{R\cos^2\delta}{R\sin^2\delta + 2h_0}, \qquad \ldots\ldots(14)$$

approximately, so that (12) becomes

$$x \leqslant \left\{ \frac{\sqrt{1 + \dfrac{4y_m}{R+h_0} \cdot \dfrac{R\cos^2\delta}{R\sin^2\delta + 2h_0}} - 1}{\dfrac{2y_m}{R+h_0}\left(\dfrac{R\cos^2\delta}{R\sin^2\delta + 2h_0}\right)} \right\} \sqrt{\frac{R+2h_0}{R\sin^2\delta + 2h_0}}. \qquad \ldots\ldots(15)$$

For the tangential ray $\delta = 0$, and expression (15) becomes

$$x \leqslant \left\{ \frac{\sqrt{1 + \dfrac{4y_m}{R+h_0} \cdot \dfrac{R}{2h_0}} - 1}{\dfrac{y_m}{R+h_0} \cdot \dfrac{R}{h_0}} \right\} \sqrt{\frac{R+2h_0}{2h_0}}. \qquad \ldots\ldots(16)$$

Since $h_0 < R$, an approximate expression for the case $\delta = 0$ is

$$x \leqslant \sqrt{\mathrm{R}} \left\{ \frac{\sqrt{2h_0 + 4y_m} - \sqrt{2h_0}}{2y_m} \right\}. \qquad \ldots\ldots(17)$$

Figures 8, 9 and 10 have been drawn to illustrate formulae (11), (13) and (14) f specified values of the layer constants y_m and h_0. From these graphs the limitin factors at any given angle of elevation can easily be determined.

§8. COMPARISON OF PRACTICAL RESULTS WITH THEORETICAL PREDICTIONS

The above method of calculating the M.U.F. as a function of distance, f specified ionospheric conditions, has formed the basis of all the M.U.F. prediction for different parts of the world, which have been issued confidentially during t war to British Service and Civil users from the Radio Research Station, Sloug We have therefore had considerable experience of the use of the method and have been able to check its general applicability.

As has been seen above, the M.U.F. is calculated by multiplying the vertica incidence critical frequency by a certain M.U.F. factor. Now, generally, t critical frequency of the *F* layer exceeds that of the *E* layer, but, on the other han the M.U.F. factor for a given value of distance *D* is greater for the *E* layer than f the *F* layer. We cannot, therefore, say that the one or the other of the two reflec ing layers always determines the M.U.F. Actually it turns out that, in wint daytime, the *F*-layer conditions control the M.U.F., whereas in summer daytin the normal *E* layer is in control. The recombination of the ionization in t *E* layer is always so rapid after sunset that practically throughout the year t night-time M.U.F. is controlled by the *F* layer.*

The success of the application of the method in ionospheric forecastin evidently depends on an accurate forecast of ionospheric conditions at the mi point of the overhead-ray trajectory. The forecasting of normal *E*-layer conditio is accurate to a fairly high degree, so regular is this layer in its behaviour. Apa from the effects of ionospheric storms, the forecasts of *F*-layer conditions, though not so reliable as in the case of the normal *E* layer, are sufficiently accurate to permit the M.U.F. forecasts to be reliable enough for practical usefulness. On the other hand, as we shall show later, the incidence of abnormal *E*-layer ionization very often permits satisfactory communication on frequencies exceeding the predicted values based on estimates of the normal *E*- and *F*-layer ionospheric conditions. Such events are found to occur most frequently in daytime during summer when, as is well known, abnormal *E*-layer ionization is most frequently experienced.

---

* Exceptions to this general rule sometimes occur during a summer night when abnorm *E* ionization may control the M.U.F. (*vide infra*, § 8 (c)).

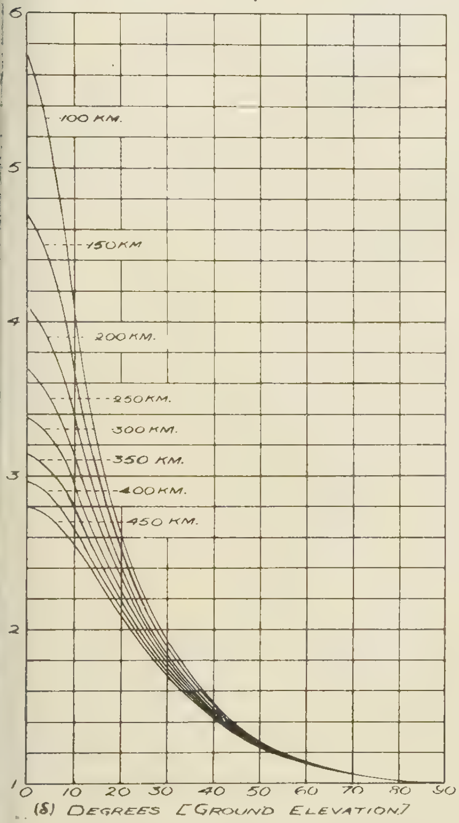


Figure 8. Maximum factor for different values of *hm*.

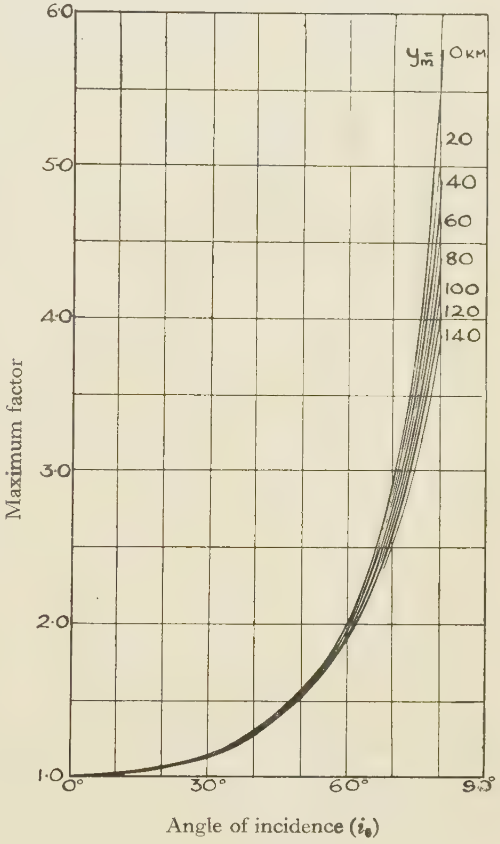


Figure 9.

Before proceeding to discuss the cases of abnormal *E*-layer transmission by way of illustrations from practical communication data, we turn to consider a number of points of physical interest in connection with the practical realization of a M.U.F. under normal conditions, such as occur when the *F* layer controls its value.

(a) *Signal intensity phenomena associated with the* M.U.F.

In making check measurements of predicted M.U.F. values from signal-intensity records of short-wave transmitting stations we have noted interesting interference phenomena as the M.U.F. conditions are approached, culminating in a very high signal value at, or near, the M.U.F. itself. All such phenomena appear to be satisfactorily explained in terms of the parabolic-layer theory of refraction outline in this paper. For convenience we shall discuss one or two theoretical points firs We shall take the case of a flat earth since it is simpler than the curved-eart treatment and illustrates all the essential physical phenomena.

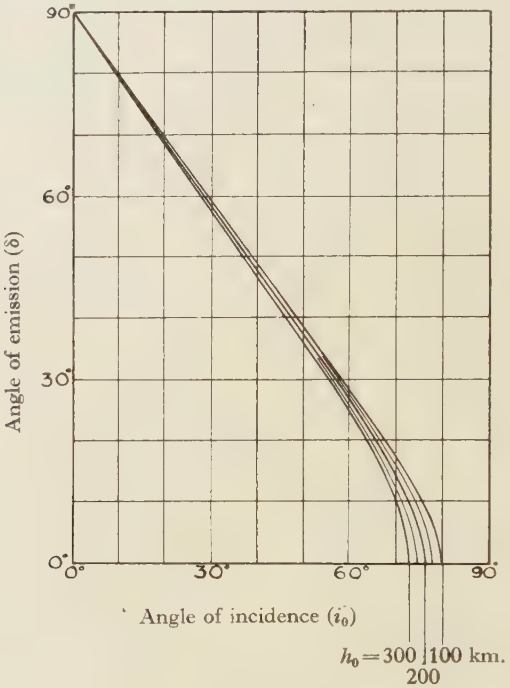


Figure 10.

In figure 11 is shown the oblique-incidence (P', f) curve calculated by our theory for a medium-distance and ordinary-ray F_2-layer transmission. It will be seen that, for operating frequencies in excess of the vertical-incidence penetration frequency $f^0_{F_2}$ there are two routes by which the waves may travel from the sender to the receiver. In the case of a frequency less than the M.U.F., say f_1, the lower equivalent path AB corresponds to a lower-ray trajectory, while the equivalent path AC corresponds to an upper-ray * trajectory. The reception of two sets of waves at the receiver causes interference phenomena if the two amplitudes are comparable.† This is especially the case as f_1 approaches the M.U.F. Now the state of interference as regards the phase difference of the two interfering components is determined by the difference of the optical paths of the upper and lower rays, and it can be shown that this quantity, for a frequency f_1, is equal to the area BNC divided by the frequency f_1. Thus, as f_1 approaches the M.U.F., the area BNC decreases and interference maxima and minima in signal strength are experienced culminating in a pronounced maximum at, or near, the exact M.U.F. value. Actually it is easy to show, by the usual method of ray tracing, that when f is equal to the M.U.F. the receiving station is situated on a caustic. The interference

---

* This upper ray is often termed the Pedersen ray. The upper and lower rays have, of course, different angles of incidence on the layer, the upper ray having the lesser angle of incidence.

† That maximum and minimum of signal amplitude are experienced just *outside* the skip zone has been noticed by Grosskopf (1940). He explains these phenomena as due to interference between the Pedersen ray and the normal ray, but gives no quantitative treatment of them.

henomena experienced at the boundary of the skipped distance have therefore a ertain resemblance to those experienced in the case of the rainbow. Just within he illuminated area there are the maxima and minima of intensity described bove, and interpreted physically as due to the simultaneous reception of the ower and upper rays, while the intensity within the shadow (" silent zone ") falls ff exponentially within the skipped distance.

In practice, the high-angle Pedersen rays are often heavily attenuated except t frequencies near the M.U.F. Furthermore, under suitable conditions, both the rdinary and extraordinary ray components may be present. A practical blique-incidence (P', f) record might thus be of the form shown in figure 12, and, n this case, interference effects between two, three or even four rays will occur. hus, referring to figure 12, as the frequency f_1 is increased to $f_1+\delta f_1$ there will be small increase in optical path difference between the two magneto-ionic compoents given by the area ABCD/f_1. Similarly, for a change from f_2 to $f_2+\delta f_2$, the hanges in optical path difference, (*a*) between the two ordinary ray components upper and lower trajectories) and (*b*) between the low-angle ordinary and extraordinary components, will be proportional to the areas EFGH and HGLM respectively. Hence, for conditions such as those represented by figure 12, we may expect the following sequence of events as the frequency in use steadily approaches the M.U.F.:

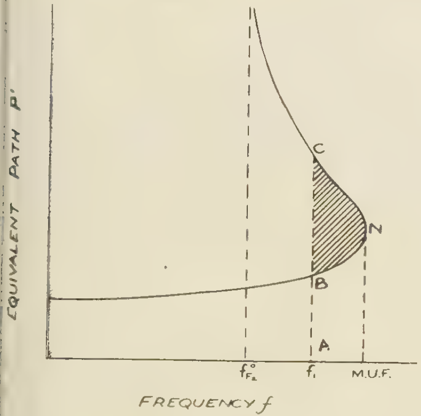


Figure 11.

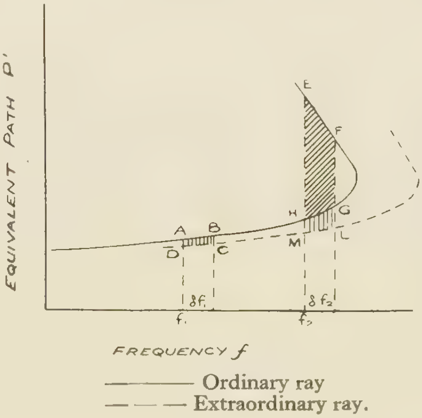


Figure 12.

(i) At first, with only a single ray present, the signal intensity may vary in an arbitrary random manner.

(ii) With the growth in amplitude of the extraordinary ray we may expect rhythmic fading of slow period, coupled with equally regular polarization changes. As the optical path difference between these two components gradually increases more rapidly (see figure 12), this rate of fading should slowly increase.

(iii) When the effect of the high angle (Pedersen) ordinary ray becomes appreciable, a rapid fading should be superposed on the slow fading.

(iv) As the M.U.F. of the ordinary component is approached, the ordinary-extraordinary fading becomes more rapid and the upper-lower ordinary ray fading slower, until the two rates may be of comparable magnitude.

(v) After the penetration of the ordinary component we are left with fair rapid fading between the two extraordinary rays (upper and lower), th rate then decreasing to zero at the M.U.F. for the extraordinary ray.

In practice it is not readily feasible to study these phenomena either b studying the signal intensity as the frequency is steadily increased up to, an beyond, the M.U.F. or, in the fixed-frequency case, by studying the signal intensit variations in space round the edge of the skipped distance. But under condition of steadily increasing (or decreasing) ionization such as occurs at sunrise (or sunset when the changing M.U.F. value passes through the fixed operating frequency used the varying features of the interference system described above appear as sequence in time.

For example, in measurements made on Oslo in the evening of 25 Januar 1946, when the *F*-layer ionization was decreasing (which would simulate th conditions illustrated in figure 12 as f_1 approached, equalled, and then exceeded the M.U.F.), the signal intensity record shown in figure 13 was obtained.

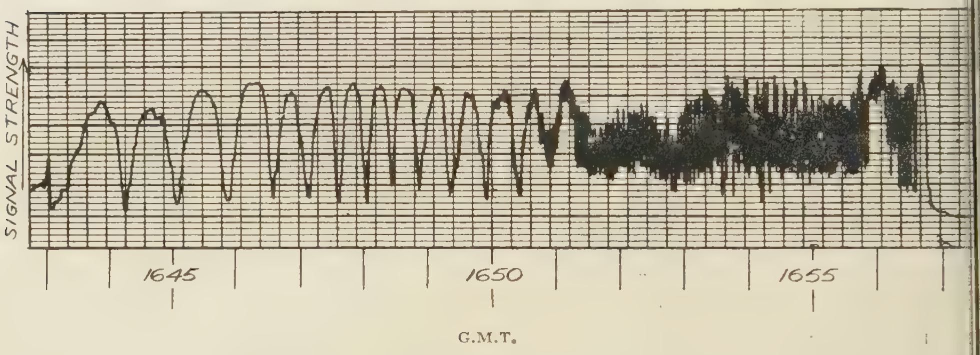


Figure 13. Variation in signal strength near M.U.F. Signal from Oslo received at Slough. Frequency 9·53 Mc./s. Distance 1140 km. 25 Jan. 1946.

Prior to 1643·5 G.M.T. the received signal intensity varied slowly in a random manner. At this time the slow rhythmic fading commenced at the rate of about 1·3 c./min. Within five minutes this rate of fading increased to about two per minute. At 1649 G.M.T. we note the first signs of the rapid fading due to the reception of the Pedersen ray; this increased in amplitude until by 1651 G.M.T. we get fast fading at the rate of about 37 cycles per minute. This then persists until the disappearance of the ordinary component at 1656 G.M.T., after which slower fading (18 c./min.), due to interference between the two extraordinary components, continues for about 40 seconds before the signal finally decreases to zero just before 1657 G.M.T. The high peaks in signal intensity which occur exactly at the maximum usable frequencies (1656 and 1656·7 G.M.T.) do not show up well on this record because of overloading of the receiver for very strong input signals. It is interesting to compare the observed rates of fading shown on this record with values calculated for a parabolic layer from normal-incidence (h', f) observations made at Slough and Burghead at the time of the field-strength record. (The mean latitude of the Slough-Oslo transmission path is approximately equal to the mean atitude of Slough and Burghead.) These calculated values are given in table 2 below.

Table 2

Frequency f (Mc./s.)	Ratio $\frac{f}{\text{M.U.F.}}$	Difference in optical path (δP)		Calculated rate of fading (cycles per minute)	
		Pedersen-low ray (km.)	Ordinary-extraordinary (km.)	Pedersen-low ray	Ordinary-extraordinary
9·53	0·91	18·0	0·1	170	1·0
	0·94	8·0	0·15	110	2·1
	0·97	2·2	0·20	60	5·5
	0·985	0·7	0·38	36	20·0

It will be seen that the calculated rates of fading agree reasonably well with the xperimental values.

b) *The variability of the F_2 layer critical frequency*

In applying normal incidence ionospheric data to communication problems t is often necessary to estimate conditions at a point many hundreds of kilometres from the nearest ionospheric laboratory. There is ample evidence to show that or small latitude changes (up to, say, 10°) the variations in the M.U.F. factor are mall compared with those in the normal incidence critical frequency of the F_2 ayer (there is no such variability in the case of the *E*-layer critical frequency). orecasts have, of course, to be made using monthly mean values and, even under uiet ionospheric conditions, we have found that the standard deviation from the monthly mean value may be of the order of 12 to 20%. (An extended analysis of his phenomenon made by Appleton and Naismith has shown that, for the latitude f Slough (51½° N.) and for years of marked solar activity, the spread of the daily alues is generally greatest for the equinox periods March and September, ndicating clearly the influence of magnetic storms.) Thus interpolated or xtrapolated values of the M.U.F. based on monthly mean values will not necessarily e really accurate for any given day. To allow for these variations it is customary, or practical purposes, to advocate operation on frequencies 15% less than the redicted value based on monthly mean values of the normal incidence critical requency.

c) *The influence of abnormal E-layer ionization on oblique incidence transmission*

There is now available a considerable body of evidence showing that satisactory radio communication can quite frequently be established on frequencies igher than the maximum usable frequencies predicted by the method described bove, which is based on the regular variations of the *E* and F_2 layers. During he war many examples of this phenomenon were brought to our attention by the British Broadcasting Corporation Engineering Division, who cited to us cases of atisfactory long-distance broadcasting on frequencies which considerably xceeded the maximum usable frequency predicted in the normal way. The information kindly furnished to us by the B.B.C. related both to "multiple-hop and "single-hop" transmission. In both cases it had been noted that discr pancies of the kind in question occurred most frequently in daytime and in summe

Now as far back as 1935 attention was drawn to the importance of abnorm *E*-layer ionization in facilitating long-distance high-frequency communicatio Appleton and Naismith (1935), for example, found that this type of ionizatio particularly in summer daytime, often returns radio energy copiously at frequenci greatly exceeding the critical frequency of the normal *E* layer. "Under suc conditions", they concluded, "the upper limit of frequency usable in long-distanc communication can be abnormally high, since the usual restriction by F_2 regio electron limitation is then not operative".

We have made a special study of the B.B.C. "single-hop" transmissio which relate to the reception of Daventry short-wave signals in Gibraltar an Algiers, the results of which appeared in a confidential paper circulated by th Radio Research Board in August 1944. Generally we found that the B.B.C results could be explained partly by reflections from the normal *E* and F_2* laye and especially by the intervention of abnormal *E*-layer ionization which provide copious reflection in summer daytime.

More recently, some particularly striking evidence relating to communicatio on frequencies exceeding the predicted values has been brought to our notice b Dr. L. P. Wheeler, Chief of the Technical Information Division of the Engineerin Department of the Federal Communications Commission, U.S.A., to whom we are indebted for permission to quote here. During the summer of 194 Dr. Wheeler found that strong signals were received in the 40 to 50 Mc./s. band at distances of the order of 1000 to 1500 km. from the sender, whereas for th particular conditions in question the maximum usable frequency for reflection by the F_2 layer or by the normal *E* layer would have been about 12 and 16 Mc./s respectively. It is clear, therefore, that abnormal *E*-layer ionization was responsible as the reflecting stratum.

Now the M.U.F. factor for the abnormal *E* layer for a distance of 1500 km. i approximately 5·0, so that for the successful transmission of, say, 44·3 Mc./s. th critical penetration frequency of the layer must have exceeded 8·9 Mc./s. So high a critical frequency is only characteristic of what Appleton and Naismith hav termed the "intense *E*-layer" condition under which the penetration frequency o the abnormal *E* layer exceeds the normal F_2 layer critical frequency.

To test whether an explanation along these lines was the correct one, we hav compared the reception results of Dr. Wheeler with the information availabl concerning the incidence of "intense *E*-layer" ionization in the U.S.A. In figure 14 are shown plotted the monthly figures of the time of occurrence o substantial reflection of 44·3 Mc./s. waves over a distance of 1428 km., together with the frequency of occurrence of normal-incidence critical frequencies of th abnormal *E* layer exceeding 9 Mc./s. at vertical incidence at Washington In each case the observations refer to the period 0600–2400 L.M.T. There ca be little doubt, from a comparison of these graphs, that the phenomena are related. An exact and detailed correspondence would not be expected in view o the fact that the observations relate to different sites, and in view of the wel nown non-uniform horizontal distribution of both abnormal *E*-layer and "intense -layer" ionization. We can, however, be quite confident that these occasional ccessful transmissions on frequencies up to twice the normally predicted values e due to the intervention of abnormally dense sporadic ionization at the level of e *E* layer.

---

* Allowance for the "longitude effect" described recently by one of us (Appleton, 1946) turned out to be particularly important in this connection.

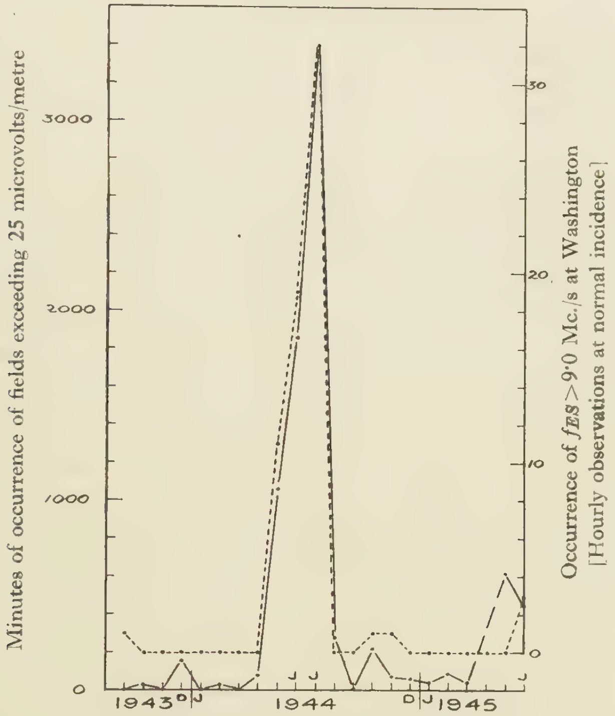


Figure 14. Propagation of 44 Mc./s. over distance of 1428 km.
----- *fES* at Washington.

APPENDIX

Note on the calculation of skipped distance for a compound ionospheric structure

For a plane ionosphere structure of the form shown in figures 3 and 4, the lation between i_0 and x for minimum D can readily be shown to be

$$\tan^2 i_0\left\{\frac{y_{\mathrm{F}} x_{\mathrm{F}}'^2}{1-x_{\mathrm{F}}'^2} - \frac{2y_{\mathrm{E}} x_{\mathrm{E}}'^2}{x_{\mathrm{E}}'^2-1} - (h_{\mathrm{F}}-2y_{\mathrm{E}})\right\} = \left\{(h_{\mathrm{F}}-2y_{\mathrm{E}}) + x_{\mathrm{E}}' y_{\mathrm{E}} \log_e\frac{x_{\mathrm{E}}'+1}{x_{\mathrm{E}}'-1} + \frac{x_{\mathrm{F}}' y_{\mathrm{F}}}{2}\log_e\frac{1+x_{\mathrm{F}}'}{1-x_{\mathrm{F}}'}\right\}, \quad \ldots\ldots(18)$$

here

$$x_{\mathrm{E}}' = \frac{f_{\mathrm{max}}}{f_{\mathrm{E}}^0}\cos i_0 \quad \text{and} \quad x_{\mathrm{F}}' = \frac{f_{\mathrm{max}}}{f_{\mathrm{F}}^0}\cos i_0.$$

or a curved ionosphere of the same structure, the angles of incidence of the waves the *E* and *F* layers are unequal. In this case the corresponding formula is milar to the above, but with additional terms in $1/R$. Such a formula cannot, owever, readily be put in an explicit form suitable for direct computation. ence, in making exact M.U.F. calculations for comparison with the results of the approximate method outlined in the main text, it becomes necessary to determin the skipped distances directly by plotting distances against angle of incidence fo given ionospheric conditions. For the curved ionosphere structure shown i figure 7 the true range D is given by

$$\left.\begin{aligned} D= & \frac{4R}{R+h_{\mathrm{E}}} \,.\, x_{\mathrm{E}} \,,\, y_{\mathrm{E}} \,.\, \sin i_{\mathrm{E}} \coth^{-1} \frac{x'_{\mathrm{E}}}{1-\dfrac{x_{\mathrm{E}}^2 y_{\mathrm{E}} \sin^2 i_{\mathrm{E}}}{R+h_{\mathrm{E}}}} \\ & + \frac{2R}{R+h_{\mathrm{F}}} \,.\, x_{\mathrm{F}} \,.\, y_{\mathrm{F}} \,.\, \sin i_{\mathrm{F}} \tanh^{-1} \frac{x'_{\mathrm{F}}}{1-\dfrac{x_{\mathrm{F}}^2 y_{\mathrm{F}} \sin^2 i_{\mathrm{F}}}{R+h_{\mathrm{F}}}} + D_2, \end{aligned}\right\} \quad \ldots\ldots(19)$$

where $D_2 = 2R(\theta_1 + \theta_2)$, θ_1 and θ_2 being given in the equations

$$\tan i_{\mathrm{E}} = \frac{R \sin \theta_1}{h_{\mathrm{E}} + R(1-\cos\theta_1)}, \qquad \tan i_{\mathrm{F}} = \frac{(R+h_{\mathrm{E}}+2y_{\mathrm{E}}) \sin\theta_2}{h_{\mathrm{F}} - \{(R+h_{\mathrm{E}}+2y_{\mathrm{E}})\cos\theta_2 - R\}}.$$

For the approximately equivalent region we should use the following formula fo calculating i_0:

$$\tan^2 i_0 \left\{ \frac{y_{\mathrm{F}} x_{\mathrm{F}}'^2}{1-x_{\mathrm{F}}'^2} - (h_{\mathrm{F}}+p) \right\} = \left\{ h_{\mathrm{F}} + p + \frac{x'_{\mathrm{F}} y_{\mathrm{F}}}{2} \log_e \frac{1+x'_{\mathrm{F}}}{1-x'_{\mathrm{F}}} \right\}, \qquad \ldots\ldots(20)$$

where $$p \simeq \tfrac{2}{3}\, y_{\mathrm{E}} (f^0_{\mathrm{E}}/f^0_{\mathrm{F}})^2.$$

The corresponding skipped distances would then be calculated in the usual wa from

$$D = \frac{2R}{R+h_{\mathrm{F}}+p} \,.\, x_{\mathrm{F}} \,.\, y_{\mathrm{F}} \,.\, \sin i_0 \tanh^{-1} \frac{x_{\mathrm{F}} \cos i_0}{1-\dfrac{x_{\mathrm{F}}^2 y_{\mathrm{F}} \sin^2 i_0}{R+h_{\mathrm{F}}+p}} + D'_2, \qquad \ldots\ldots(21)$$

where $D'_2 = 2R\theta'$ and θ' is given in the equation

$$\tan i_0 = \frac{R \sin\theta'}{h_{\mathrm{F}} + p + R(1-\cos\theta')}.$$

ACKNOWLEDGMENT

The work described above was carried out as part of the programme of th Radio Research Board, and this paper is published by permission of the Departmen of Scientific and Industrial Research.

REFERENCES

Appleton, E. V., 1939. "The Structure of the Atmosphere as deduced from Ionospheri Observations." *Quart. J.R. Met. Soc.*, **65,** 324.

Appleton, E. V., 1946. "Two Anomalies in the Ionosphere." *Nature, Lond.*, **157,** 691

Appleton, E. V. and Beynon, W. J. G., 1940. "The Application of Ionospheric Data t Radio Communication Problems", Part I. *Proc. Phys. Soc.*, **52,** 518.

Appleton, E. V. and Naismith, R., 1935. "Some Further Measurements of Uppe Atmospheric Ionization." *Proc. Roy. Soc.*, A, 871, **150,** 685.

Grosskopf, J., 1940. *Telegr. Fernspr.-Funk.- u. Fernsehtech.*, **29,** 127.

Martyn, D. F., 1935. "The Propagation of Medium Radio Waves in the Ionosphere." *Proc. Phys. Soc.*, **47,** 323.

Millington, G., 1938. "The Relation between Ionospheric Transmission Phenomena at Oblique Incidence and those at Vertical Incidence." *Proc. Phys. Soc.*, **50,** 801

Smith, N., 1937. "The Extension of Normal Incidence Ionosphere Measurements t Oblique Incidence Radio Transmission." *J. Res. Nat. Bur. Stand., Wash.*, **19,** 89

AN INFRA-RED SPECTROSCOPE WITH CATHODE-RAY PRESENTATION

By E. F. DALY and G. B. B. M. SUTHERLAND
Pembroke College, Cambridge

MS. received 30 *August* 1946

ABSTRACT. An instrument developed for the visual presentation of infra-red spectra on a cathode-ray screen is described. The radiation detector employed is a high-speed bolometer made by the Bell Telephone Laboratories having a time constant of less than 0·01 second. The performance of the instrument is such that a range of 2·5 to 3·5 μ, anywhere between 1 and 16 μ, can be scanned in 14 seconds. The cathode-ray screen used has very long persistence, so that a steady "picture" is obtained. The resolving power at these speeds is sufficiently good for most work on complex molecules. Illustrations of typical spectra are given and proposals are outlined for further improvement and development.

---

§1. INTRODUCTION

In recent years it has become very important to improve the speed with which absorption spectra may be plotted in the infra-red. The aim has been to make the recording of spectra a routine matter so that attention can be concentrated on the correlation and interpretation of data on series of molecules of related constitution. In this way bands which characterize particular chemical bonds and groupings can be identified with certainty and used as guides in elucidating the structure of molecules of unknown constitution. Several recording spectrometers have been described which plot absorption spectra between 2 and 15 μ (with moderate resolving power) in a time of the order of half an hour. (For example, Barnes *et al.*, 1945; Brattain and Beeck, 1942; McAllister *et al.*, 1941; Sutherland and Thompson, 1945; Wright, 1941.) However, in order to exploit fully the infra-red method of analysis, it is desirable to be able to see the infra-red spectrum, as this enables one to apply the method to problems in which the spectrum is changing rapidly, e.g. in chemical reactions, in changes of state and in other transitory phenomena. Quite apart from making these new fields accessible, such an infra-red spectroscope has great advantages as an inspection instrument and for doing a quick survey of a problem in order to enable one to decide whether a more detailed study is worth while, using a conventional recording instrument.

Before the instrument described here was commenced there had been only one attempt (Baker and Robb, 1943) to make an infra-red spectroscope using cathode-ray presentation. The principal controlling factor in such an instrument is the speed of the detector system. In the apparatus of Baker and Robb this consisted of a bolometer and a Moll micro-galvanometer, having a combined time constant of 0·2 sec. With this they could arrange to display 75 points in a spectrum in a time of one minute, i.e. about 1 point a second. The resulting picture must either cover an inconveniently small portion of the infra-r spectrum rather slowly, or a reasonable range with unsatisfactory resolvin power. Unfortunately no absorption spectra were given by Baker and Rob from which the actual performance could be judged. In our instrument w employ as radiation detector a Bell Telephone Laboratories thermistor bolomet with a time constant of just under 0·01 sec. (Brattain and Becker, 1946). It operated by radiation interrupted at 20 c./s. and is followed by an amplifi responding to the 20 c./s. output voltage of the detector, but not to slow voltag changes, or "drift". With this system we are able to scan a range of $2{\cdot}5\,\mu$ $3{\cdot}5\,\mu$ between $1\,\mu$ and $16\,\mu$ in a time of 14 sec. and to obtain at this scanning spee a resolving power which is sufficiently good for most work on complex molecule Preliminary descriptions of an apparatus have already appeared (Daly an Sutherland, 1946).

§2. SPECTROMETER AND DETECTOR

A Nernst glower was used as a source of infra-red radiation, without speci precautions other than shielding from draughts and the use of baretter lamp to drop the mains supply voltage. The radiation is collected by a mirror an focused on the spectrometer entrance slit through a rotating sectored disc which gives a radiation intensity modulated in square wave form at about 20 c./s The frequency is controlled by a mechanical governor and set by means of stroboscope disc and neon lamp fed from the 50 c./s. mains. It may be varie over a range of 10 c./s. to 25 c./s., but it is found that certain sub-multiples of th mains frequency are to be avoided, as these may give rise to slow beats wit residual mains hum in the amplifier. A frequency of exactly 20 c./s. has been found quite suitable for normal operation.

The spectrometer used was an old Bellingham and Stanley instrumen having Littrow mounting of a 30° back-silvered rock-salt prism 5 cm. high and with a 3 cm. base. It has an aperture ratio of approximately F/12, which is rather low for this purpose, but was the only one readily available. The prism table drive has been modified (see figure 1) to operate from a continuously-rotating cam. This cam provides a uniform traverse of the prism table through an angle of about 1° for 300° rotation of the camshaft and a rapid flyback to the starting point over the remaining 60° of rotation. A screw control moves the cam roller relative to the prism table arm, thus giving a means of setting the start of the scan to any given wave-length. The angular traverse of the prism table during one scan may be varied by employing a selection of cams, but in our first model we have found that a cam cut to give a 300 cm^{-1} scan in the $7\,\mu$ region is most generally useful.

The scan repetition frequency used is 3·5 per minute, which is well within the persistence limit of the cathode-ray-tube screen. The scanning speed is thus about 20 cm^{-1}/sec. in the $7\,\mu$ region. For the present spectrometer and detector system, this appears to be the most satisfactory compromise. It is hoped that this may be raised to 100 cm^{-1}/sec. with similar resolving power by the use of a spectrometer with larger aperture.

As mentioned above, the detector used is a thermistor bolometer developed by the Bell Telephone Laboratories. It consists of a flake of thermistor material ith suitable current leads mounted on glass backing and having a receiver ea of 2 mm. × 0·2 mm. The receiver operates in air but is sealed off from ie atmosphere, radiation being admitted through a rocksalt window. This iermistor element is used in series with a compensating element matched it in resistance/current characteristics, but not exposed to radiation. The vo elements form the arms of a bridge (see figure 2) and from their junction ie output signal is fed via a condenser to the grid of the first valve of the head nplifier.

If the radiation falling on the detector is modulated in this way, and the head nplifier responds to the modulation frequency but not to frequencies lower ian a few cycles per second, the amplifier output will depend only on the nplitude of the *modulated* radiation and not on relatively slower radiation ianges, or drift, caused by varying ambient temperature in any part of the system fter the chopper. Sudden changes in radiation intensity, such as the switching n or off of lights, produce transient responses, which, however, are readily ecognizable. In our system a modulation frequency of 20 cycles per second s used, a value which permits maximum signal-to-noise ratio to be reached, ut which allows the amplifier pass-band to exclude interference from 50 c./s. nains hum.

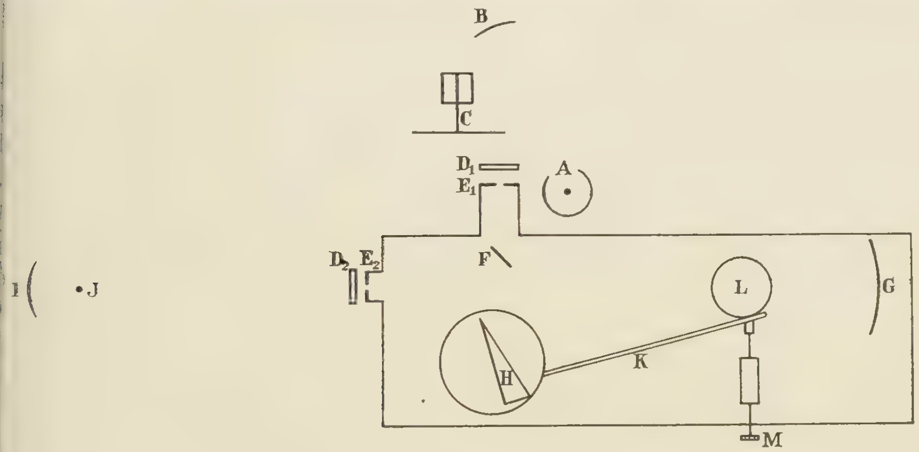


Figure 1. General lay-out of spectrometer.

A	Nernst filament.	E_1	Entrance slit.	I	Condensing mirror.
B	Condensing mirror.	E_2	Exit slit.	J	Bolometer.
C	Chopping disk.	F	Plane mirror.	K	Prism table arm.
1, D_2	Alternative positions of absorbing material.	G	Collimating mirror.	L	Cam.
		H	Prism.	M	Control for cam setting.

The thermistor bolometer is operated with a d.c. bias voltage of between 100 nd 150 across each element. Sensitivity to radiation increases with bias voltage, ut so also does current noise in the thermistor flake. In addition, the flake esistance falls appreciably with increasing operating temperature, thus causing fall in Johnson noise with increasing bias. A compromise must therefore e effected by choosing a bias voltage which is sufficiently high to give good ensitivity, but not so high that the current noise in the detector greatly exceeds he Johnson noise. It is found in practice that the curve of signal-to-noise atio against bias voltage is fairly flat over the bias range 100–150 volts. Below this it falls off, and bias above this range tends to drive the bolometer into a unstable condition. Under optimum bias conditions, the resistance of th bolometer element is of the order of 2 megohms, so that, for noise purpose the output signal is generated across a resistance of a megohm. The voltag sensitivity is approximately 250 μv. per μw. of radiation, interrupted at 20 c./s for 100 volts D.C. bias per element.

§ 3. AMPLIFICATION AND DISPLAY

The present head amplifier (figure 2) was designed after some experienc with that used in the original system (Daly and Sutherland, 1946). It has th following characteristics:—60 db. gain at 20 c. s., a response flat within 0·5 d over the range 2 c./s. to 40 c./s., a noise level corresponding to a noise resistanc of 100 000 to 200 000 ohms at the first grid, and an input impedance of not le than 5 megohms at 20 c./s.

Those features of the amplifier concerned with noise level will be discusse first. Thermal and current noise arising in the thermistor elements must b regarded as unavoidable. The remaining major sources of noise are (i) fluctuatio in the power supplies to the bolometer bridge and to the amplifier, (ii) leakag through the coupling condenser between the thermistor bridge and the gri of the first amplifier valve, (iii) flicker noise in the early stages of the amplifie Fluctuations in the power supplies may be reduced to a low value by adequat stabilizing and filtering circuits. The H.T. supplies are stabilized by th customary hard-valve series stabilizing circuit (see figure 2), the output fro which is fed to a chain of gas-discharge valves which provides tappings at −15 0, +150 and +300 volts. The −150, +150 and +300 volt supplies are furthe smoothed by choke-condenser filters and by the decoupling networks in the hea amplifier itself. The second source of noise was avoided by operating th thermistor bolometer bridge from the −150 and +150 volt lines. Arrangement are made for varying the bias while keeping the bridge mid-point approximatel at earth potential. In this way the first coupling condenser (which is selecte for low leakage) is not subjected to potential difference of more than a few volts

Flicker noise in the valves is a well-known source of difficulty in designin low frequency amplifiers. The 6SJ7 valves used have relatively low flicke noise, and were operated at low screen voltages (about 10 v.) and with anod voltages of about 40 in the earlier stages to minimize ionization of any residua gas. It was not considered profitable to operate the valves at reduced heate voltage, as is often recommended, since this may shorten the useful life of th valves and was not found to give a significant improvement in signal-to-nois ratio. Carbon resistors carrying appreciable current were avoided wher possible in the earlier stages of the amplifier. Hum from the 50 c./s. main was reduced by employing a smoothed d.c. heater supply, and by complet electrostatic shielding of the bridge and first two amplifier stages.

Since, in the thermistor bridge, the signal voltage is generated across an effective impedance of a megohm, it is desirable that the input impedance of th amplifier should be as large as possible compared with this if loss of signal is t be avoided. This high input impedance was secured by using a cathode followe as the first stage, the cathode load being common to the first two valves. Sinc

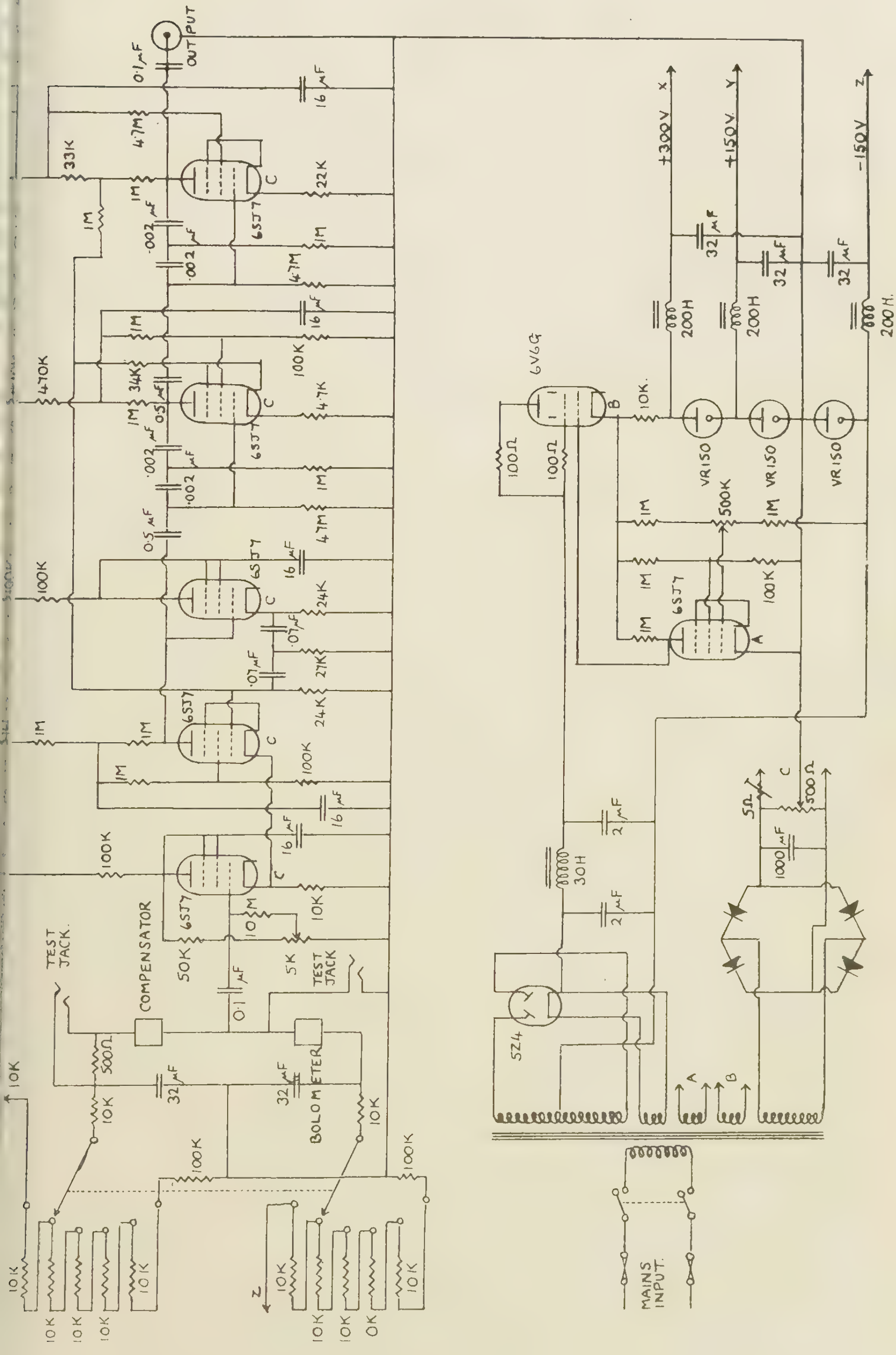


Figure 2. Head amplifier circuit.

negative feedback is applied to the control grid of the second stage, the cathod input impedance is relatively high, and the operation of the first valve is th not impaired.

The response of the head amplifier was made wide in order that each energ pulse on the detector should give rise to a corresponding discrete voltage puls at the amplifier output. Although in the circuits described below which follow the head amplifier, this band-width is greatly reduced by smoothing, it has been maintained at 40 c./s. up to detector level, as we intend to introduce shortly system analogous to that of Baker and Robb (1943) for producing a direct plo of percentage transmission against wave-length instead of an energy trace. Thi system may involve the transmission of alternate pulses of widely differen amplitudes, for the accurate reproduction of which a wide response band i essential.

Figure 3 shows the shape of the head amplifier response. This type o response was obtained by the use of negative feedback, via a high-pass T-section filter, between anode and control grid of each pentode amplifier stage (see figure 2) and by overall negative feedback from the anode of the final stage to the control grid of the first pentode stage, and from the cathode of the second pentode stage also to the control grid of the first. In spite of the multiplicity of feedback loops, the amplifier proved to be quite stable.

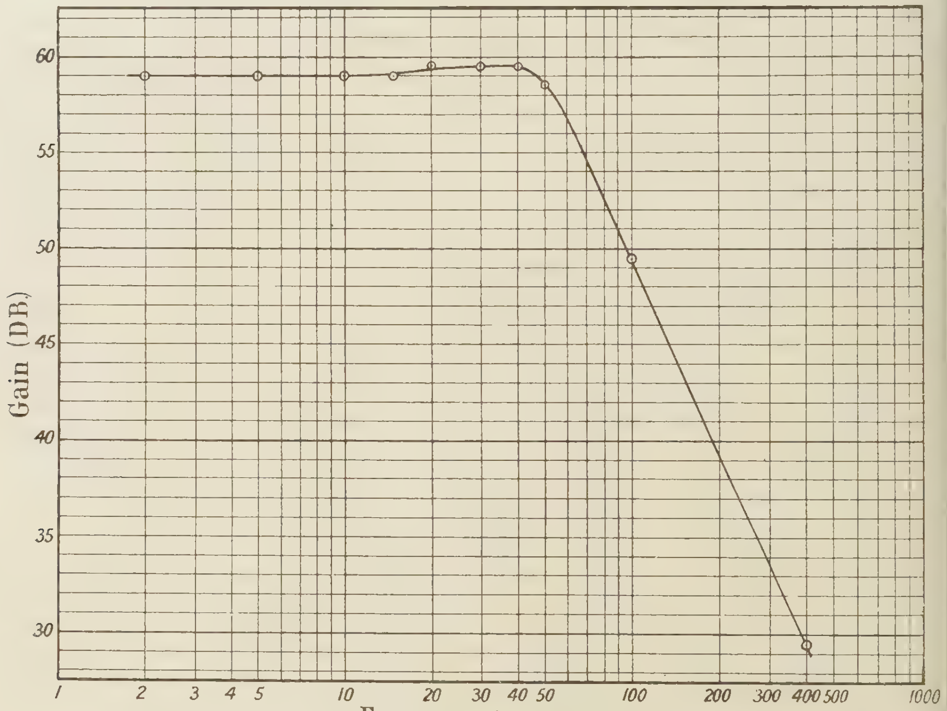


Figure 3. Head amplifier response.

The detector and head amplifier have been considered together as one unit— i.e. a radiation detector, whose output is at a sufficiently high level for it to be unaffected by the presence of normal noise in the succeeding electronic circuits. This unit has been arranged, as mentioned above, to give a close correspondence between input radiation pulses and output voltage pulses.

If following circuits are such that this correspondence is maintained up to the cathode-ray tube used for display, each radiation pulse will be represented by a proportional deflection (see plate 1 (*a*)). Alternatively, the radiation-chopping frequency may be treated as a carrier and detected, to yield finally a smoothed signal at a considerably narrower band-width (see plate 1 (*b*)). The former system is preferable where radiation energy is plentiful, and where it is desired to operate at high scanning speeds with low distortion of the form of the observed spectra. When energy is strictly limited, however, the signal-to-noise ratio at the display end may be increased by detecting and smoothing the interruption-frequency signal to a band-width of the order of 1/10 to 1/20 of that passed by the head amplifier. This considerable reduction in band-width is necessary if discrimination against ripple at the interruption frequency is to be achieved without a very elaborate filter system.

We have developed the second system, which is convenient for the qualitative examination of spectra, in the following manner: Output signals from the head amplifier are fed through a gain control, having ten 6 db. steps, to an amplifier with 40 db. gain and a flat response from 2 c./s. to 40 c./s., which is similar in design to the head amplifier. The output from this second amplifier is detected by a diode and the output smoothed by means of a three-section, low pass, resistance-capacity filter. The 20 c./s. ripple is reduced in amplitude to 1 to 2% of the signal. The output of this filter, which has a band-width of about 2 c./s. is used to control the amplitude of a 500 c./s. square wave carrier (see figure 4). The circuit may be so adjusted that zero energy falling on the detector gives a very small carrier amplitude.

This 500 c./s. signal is then amplified by one stage with continuously variable gain (max. 35 db.) and finally by a paraphase amplifier in which the two anodes feed the Y plates of the cathode-ray tube. Between the variable gain stage and the grid of the paraphase amplifier, there is interposed either a d.c. restoring diode, or a clipping circuit with adjustable bias. The latter circuit is used if it is desired to have a sharp base-line which can be set precisely to zero energy on the detector, even in the presence of appreciable noise background. These circuits are shown in figure 4. An advantage of the 500 c./s. square wave carrier system is that the C.R.T. spot remains stationary for a considerable fraction of each cycle, so giving bright upper and lower limits to the pattern (see plate 1 (*b*)).

The cathode-ray tube is a 12-inch electrostatic deflection type, with a double screen of the type used in radar P.P.I. display units. A spot of the order of 0·5 mm. diameter is used and the maximum dimensions of the spectrum pattern are 15 cm. by 15 cm.

A time-base is provided by an accurate 100,000 ohm potentiometer linked to the spectrometer prism table drive. The potentiometer slider is connected to one X plate, and the other X plate to a voltage corresponding to the potentiometer mid-point. Continuous rotation of the potentiometer is possible, but for 60° the slider is not in contact with the winding. This 60° interval is used for the return of the prism table to its starting point, and during it the cathode-ray-tube trace is suppressed by a flasher and relay system driven from the sam shaft as the time-base potentiometer and operating a brightness control.

§4. ILLUSTRATION OF SPECTRA

In plate 1 we give a comparison between the two methods of display. I the upper photograph we have the trace produced by the wide-band (40 c./s. undistorted signal from the head amplifier. This is the same type of displa as in our preliminary note in *Nature* (Daly and Sutherland, 1946). In th lower photograph we have the trace due to the narrow-band (2 c./s.) smoothe and re-modulated signal. The spectral region covered (roughly 1–4 μ) and th radiation intensity at the detector are the same for both plots. The sourc of radiation was a Nernst filament which has its maximum near 2 μ; superimposed atmospheric absorption bands at 1·4, 1·9 and 2·8 μ are clearly visible. In the upper photograph more detail is visible and this detail is reproducible with great accuracy, but from a psychological point of view the lower trace is preferable. While the paper was being written a note appeared (King *et al.*, 1946) giving an account of a copy of our original apparatus but with the introduction of smoothing. The authors of that note appear to consider that smoothing is always an advantage. They claim that it makes the detection of weak absorption bands easier. We would point out that this is not so, and *for the same scanning speed* smoothing must always involve some sacrifice of information. However,

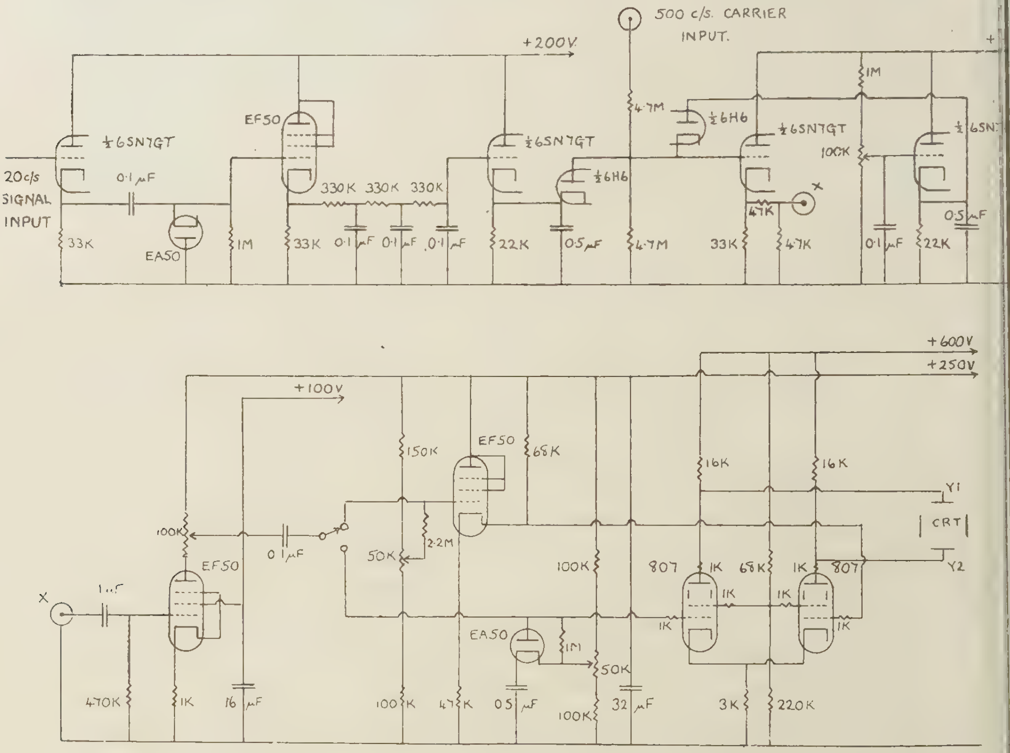
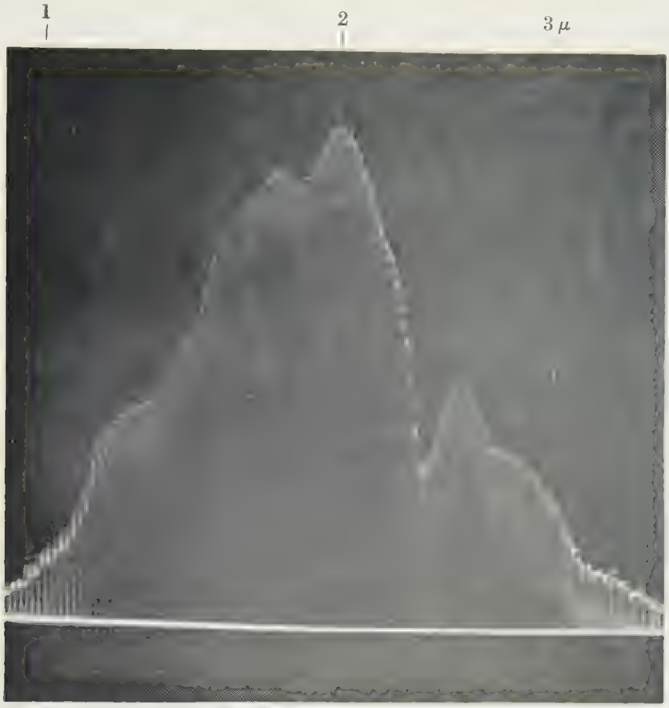
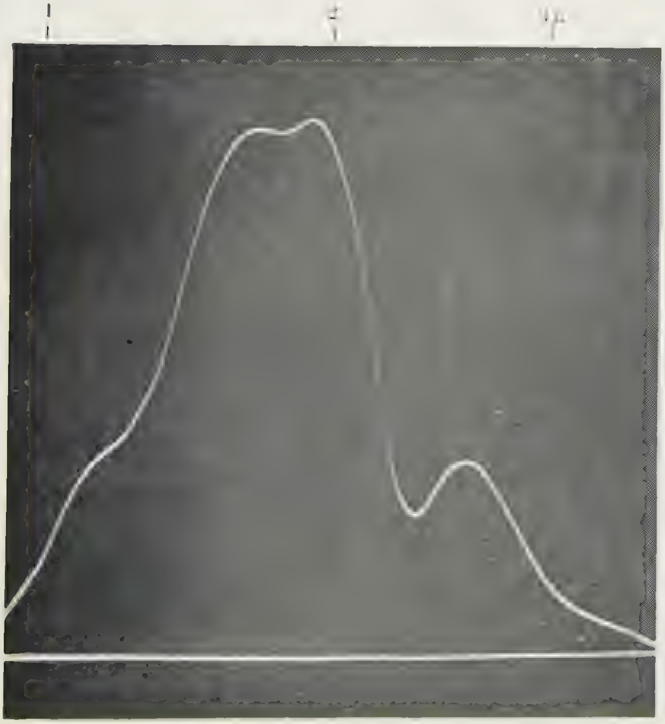


Figure 4. Circuits for smoothing, remodulation, and for final paraphase amplifier.



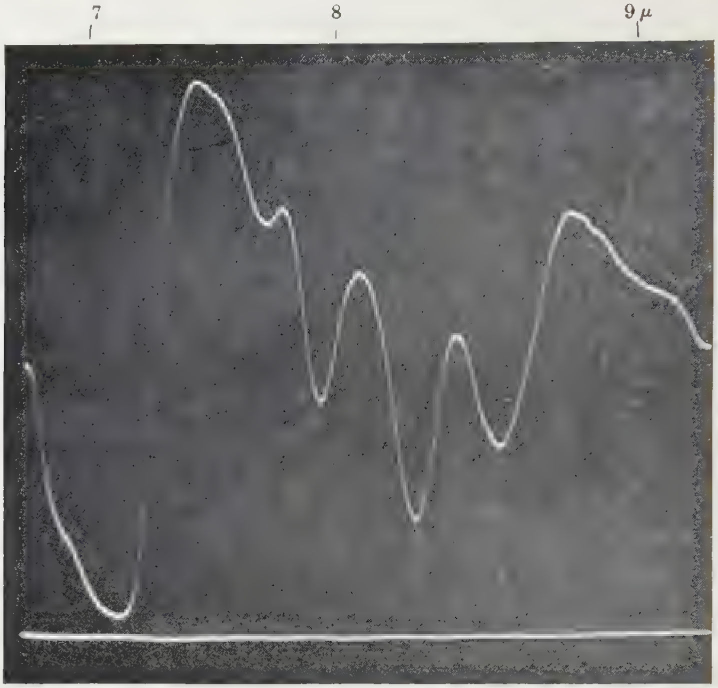
(a)



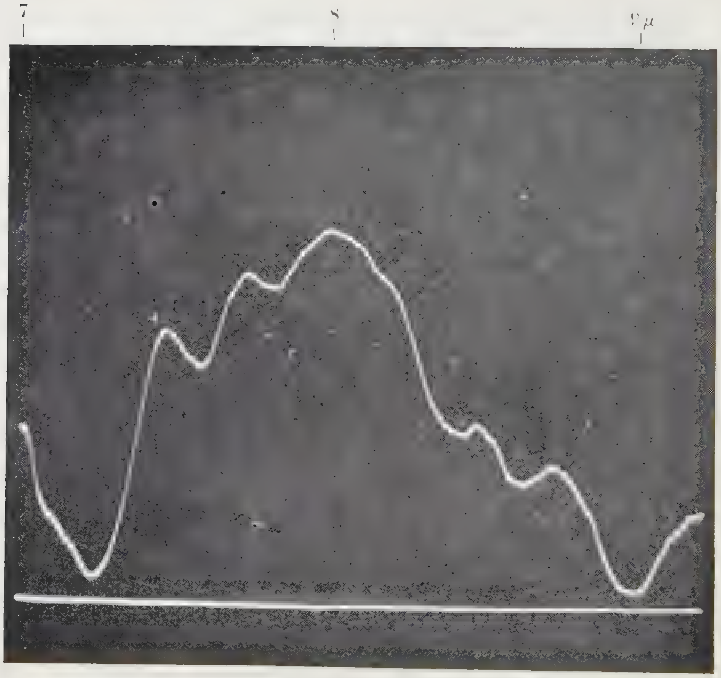
(b)

Plate 1. Emission spectrum from a Nernst filament.

To face page 84



(a) 2.2.4 Trimethylpentane.



(b) 2.3.4 Trimethylpentane.

Plate 2. Absorption spectra of isomeric hydrocarbons.

For certain purposes this sacrifice may be unimportant compared to the psychological gain, and it is for this reason that we have developed both systems, which can be used interchangeably.

In plate 2 we illustrate the use of this instrument for rapid identification and analysis of chemicals. The upper photograph gives the absorption spectrum of 2.2.4 trimethylpentane over the approximate range 7–9 μ. The source of radiation was again a Nernst filament; the absorption minima in the energy curve are at 7·3, 7·8, 8·0, 8·3 and 8·5 μ and completely characterize this hydrocarbon. For comparison we give in the lower photograph the corresponding spectrum of the isomer 2.3.4 trimethylpentane with its equally characteristic series of bands at 7·3, 7·6, 7·9, 8·45, 8·6 and 8·9 μ. When one considers that these two plots were obtained in a time of approximately 30 seconds, and that there is every likelihood that this time can be reduced by a factor of 5 in the very near future, the potentialities of this instrument as a tool for chemical research and analysis need no further emphasis.

We have already used the spectroscope to follow dissociation phenomena and are at present studying certain chemical reactions. An account of this work will be more appropriately given at a later date in separate papers.

§ 5. FUTURE DEVELOPMENTS

In constructing this apparatus we have concentrated on simplicity and essentials. We considered the important thing was to show the practicability of such a spectroscope rather than to develop a very complicated instrument with elaborate refinements. Since an instrument of this type will have many diverse applications, various forms will have to be developed for special purposes. However, there are certain features to which we wish to draw attention, as we believe them to be of general interest to all future users.

Let us consider first the optical and mechanical parts of the spectroscope. Since in the infra-red the principal limitation to resolving power is determined by the energy passing through the exit slit, we must strive to maximize this energy for a given frequency range in the spectrum. This can be done by increasing the size of the prism, the aperture ratio and the length of the entrance slit.* Our present instrument employs a 30° prism with 3 cm. base with an effective aperture ratio F/12. We propose to use in future a 60° prism with 7 cm. base and an aperture ratio F/4·5. The limitations imposed here are solely the size of prism available and the cost of appropriate mirrors of the requisite focal length and optical accuracy. If the mirror is used off-axis, this last may be a serious consideration and it would be advisable to consider using an optical scheme such as that due to Pfund (1927) in which the collimating mirror is used on-axis.

In the present instrument we have used a single cam and a single scanning speed. The cam could be adjusted so that the 1° movement of the prism could take place in any part of the spectrum. This has the advantage of continuous variability, but is unsuitable in that it is designed really for the 6–8 μ region of the spectrum. Thus, at shorter wave-lengths, the spectrum is too crowde while at longer wave-lengths it is too spread out. An alternative method wou be to have a series of, say, 5 cams to cover the range 1–16 μ with some overl from one cam to the next. These cams could be interchangeable by mea of a simple external control and calibrated to give a linear frequency scan with their respective ranges. Our original system has the great advantage that i flexibility allows more latitude in the comparison of absorption bands whic might otherwise fall on different cams and we are retaining it with modificatio which allow us to vary independently (1) the starting point of the scan, (2) t width of spectrum being scanned, and (3) the speed of scanning. At prese no calibration marks are introduced in the spectrum displayed on the cathod ray tube. A number of different methods of calibration are possible, but w propose to introduce one in which the trace is darkened momentarily as t spectrometer wave-length setting passes certain fixed values.

---

* We hope to publish a full analysis of these and other factors controlling the performance of this spectroscope in the very near future.

In the investigation of absorption spectra, the direct presentation of percenta transmission *v.* wave-length is desirable. The inclusion of this feature wa decided upon at an early stage, and, as implied above, the system has been designe so that it may be added. The equipment now being set up for this purpos is in principle analogous to that of Baker and Robb (1943), but it will be see that, in view of our higher interruption frequency and potentially higher scannin speed, it is desirable to substitute a purely electronic system for their electr mechanical system of automatic gain control.

During the past six years a number of novel fast detectors have been develope for war purposes, in addition to the thermistor bolometer which we have use Some of these, such as the superconducting bolometer (Andrews *et al.*, 1946 air cell detector (Weiss, 1946) and lead sulphide cell (Cashman, 1946; Starkiewic *et al.*, 1946), while not so convenient as the thermistor bolometer, offer con siderable improvement in speed and sensitivity. Some of these are now unde investigation by us.

Although the most striking feature of the present apparatus is that it give a steady visible picture of the infra-red spectrum, its applicability to variou problems will be restricted unless a record of the spectrum can be made, a desired, in tandem with its display. The obvious method of photography i rather clumsy and slow, so we are making arrangements to use a method of ultra high-speed pen recording.

ACKNOWLEDGMENTS

This work forms part of a programme of fundamental research sponsore by the Hydrocarbon Research Group of the Institute of Petroleum. We wis also to express our thanks to the Ministry of Aircraft Production and to th Admiralty, the former for providing the thermistor bolometer and P.P.I. displa tube, the latter for allowing one of us (E.F.D.) to work on this project whil still in Admiralty employ (Oct.–Dec. 1945).

It is a pleasure to acknowledge the benefit of many stimulating discussions with Mr. Peter Fellgett, who also gave valuable assistance in the preliminary stages of this project.

REFERENCES

Andrews, D. H., Milton, R. M. and de Sorbo, W., 1946. *J. Opt. Soc. Amer.*, **36**. 353.
Baker, E. B. and Robb, C. D., 1943. *Rev. Sci. Instrum.*, **14**, 362.
Barnes, R. B., McDonald, R. S., Williams, V. Z. and Kinnaird, R. F , 1945. *J. Appl. Phys.*, **16**, 77.
Brattain, R. R. and Beeck, O., 1942. *J. Appl. Phys.*, **13**, 699.
Brattain, W. H. and Becker, J. A., 1946. *J. Opt. Soc. Amer.*, **36**, 354.
Cashman, R. J., 1946. *J. Opt. Soc. Amer.*, **36**, 356.
Daly, E. F. and Sutherland, G. B. B. M., 1946. *Nature, Lond.*, **157**, 547; *ibid.*. **158**, 591.
King, J., Temple, R. B. and Thompson, H. W., 1946. *Nature, Lond.*, **158**, 196.
McAllister, E. D., Matheson, G. L. and Sweeney, W. J., 1941. *Rev. Sci. Instrum.*, **12**, 314.
Pfund, A. H., 1927. *J. Opt. Soc. Amer.*, **14**, 337.
Starkiewicz, J., Sosnowski, L. and Simpson, O., 1946. *Nature, Lond.*, **158**, 28.
Sutherland, G. B. B. M. and Thompson, H. W., 1945. *Trans. Faraday Soc.*, **41**, 174.
Weiss, R. A., 1946. *J. Opt. Soc. Amer.*, **36**, 356.
Wright, Norman, 1941. *Ind. Eng. Chem.*, **13**, 1.

EQUIVALENT PATH AND ABSORPTION IN AN IONOSPHERIC REGION

By J. C. JAEGER,
The University of Tasmania

MS. received 24 *August* 1946

ABSTRACT. Numerical values for the equivalent path and absorption of a wireless wave incident vertically on a Chapman region are given for the cases of penetration and of reflection from either side of the region. The values are compared with those for the commonly assumed parabolic distribution of ionization. The application of the results to absorption measurements and to the reduction of experimental P', f curves is considered.

§1. INTRODUCTION

It is often convenient to be able to compare the relationship between equivalent path and frequency found experimentally, with the values calculated assuming a definite form for the distribution of ionization. For instance, Appleton (1937) has sometimes assumed that the ionization is parabolically distributed.

Chapman (1931) has made very general calculations of the distribution of ionization produced in the atmosphere by the incidence of monochromatic radiation from the sun. These are based on the assumption that the gaseous component responsible for absorbing the radiation is distributed exponentially with height, and allowance is made for the relative movement of the sun and the earth. It is shown here that the equivalent path and absorption of waves can be calculated for a region in which the ionization is distributed as indicated by Chapman's theory: such a region will be referred to below as a "Chapman region".

In studying absorption, Best and Ratcliffe (1928) have calculated values for a Chapman region, limiting the calculation to the case of waves of frequency much greater than the critical frequency for the region, so that it can be assumed that its refractive index is nearly unity. This limitation is not imposed here.

§2. DISTRIBUTION OF IONIZATION

The formulae relating ion production in the atmosphere to the mass absorptio of solar radiation were first given by Lenard in 1911. In 1931 the complete theor of the region formation as a function of the position of the sun relative to the eart was given by Chapman (1931). He shows that if monochromatic radiation fror the sun, of a wave-length that will ionize one of the atmospheric components, incident upon the earth's atmosphere, the rate of ion production q at any height above the earth's surface will be related to the zenith angle χ of the sun by th formula

$$q = q_0 \exp(1 - z - \sec\chi\, e^{-z}), \qquad \ldots\ldots(1)$$

where the earth is regarded as flat, and

$$z = \frac{h - h_0}{H}. \qquad \ldots\ldots(2)$$

Here z is the height above a datum level h_0 measured in terms of the scal height

$$H = \frac{\mathbf{k}T}{mg}, \qquad \ldots\ldots(3)$$

where $\mathbf{k}$ is Boltzmann's constant, T is the absolute temperature, g is the acceleratio of gravity, and m is the mean mass of the atmospheric molecules involved.

If it is assumed that the ions formed by the action of sunlight disappear b recombination at the rate αN^2 per c.c. per sec., where N is the number of ion existing per c.c. at any time, then N is determined during the day by the equation

$$\frac{dN}{dt} = q - \alpha N^2. \qquad \ldots\ldots(4)$$

It is known that dN/dt is small in the E-region in daylight; assuming as a first approximation that it vanishes, we have from (1) and (4)

$$N = \sqrt{(q/\alpha)} = N_0 \exp \tfrac{1}{2}(1 - z - \sec\chi\, e^{-z}), \qquad \ldots\ldots(5)$$

where $N_0 = \sqrt{(q_0/\alpha)}$.

The value of N given by (5) has a maximum

$$N_{\max} = N_0 \cos^{\frac{1}{2}}\chi \qquad \ldots\ldots(6)$$

at the level given by

$$e^{-z} \sec\chi = 1.$$

The value $\sec\chi = 1$ occurs at the equator at noon at the equinox. Thus N_0 is the maximum value of N at this place and time, and this value occurs when $z = 0$, that is, at a height h_0 above the earth's surface.

§3. ABSORPTION

When a radio wave is sent upwards at vertical incidence towards such an ionized region it is known from the theory of Lorentz that the refractive index μ and the absorption coefficient κ of the medium are given by

$$\mu^2 = 1 - \frac{N\epsilon^2}{\pi m f^2}, \qquad \ldots\ldots(7)$$

$$\kappa = \frac{\nu}{2\mathbf{c}} \frac{1 - \mu^2}{\mu}, \qquad \ldots\ldots(8)$$

where f is the frequency of the wave, ϵ and m are the charge and mass of the ions, c is the velocity of light, and ν is the collision frequency of the ions with neutral molecules. It is assumed that $(2\pi f)^2 \gg \nu^2$.

At vertical incidence the wave will be reflected from the region if there are sufficient ions at any level to reduce the value of μ to zero. This reflection occurs at the level at which $N = \pi m f^2/\epsilon^2$: if the effective ions are electrons, as is believed to be the case in E-region, this is equal to $1{\cdot}24 \times 10^4 f^2$, where f is the frequency in Mc./sec.

The value of f for which reflection occurs where $N = N_{\max}$ is called the critical frequency f_c for the region. It follows that

$$f_c^2 = \frac{\epsilon^2 N_{\max}}{\pi m} = \frac{\epsilon^2 N_0}{\pi m} \cos^{\frac{1}{2}}\chi. \qquad \ldots\ldots(9)$$

A small increase of f over f_c will cause the wave to penetrate the region.

The absorption coefficient κ involves the collision frequency ν of the electrons with neutral molecules. It is customary to assume that ν varies exponentially with height, so that

$$\nu = \nu_0 e^{-z}, \qquad \ldots\ldots(10)$$

where ν_0 is the value of ν at $z = 0$, that is at height h_0.

It is now desired to show how the total absorption varies with frequency and with the zenith angle of the sun for waves reflected from, or transmitted by, such a region.

The total absorption is $\int \kappa ds$, where the integral is taken along the whole trajectory of the wave. Thus from (2), (5), (7), (8), (10) we find for vertical incidence

$$\int \kappa ds = \int \frac{\nu}{2\mathbf{c}} \frac{1-\mu^2}{\mu} ds$$

$$= \frac{\nu_0 f_0^2 H}{2\mathbf{c} f^2} \int \frac{\exp \frac{1}{2}(1 - 3z - \sec\chi e^{-z})}{\sqrt{\{1 - (f_0^2/f^2) \exp \frac{1}{2}(1 - z - \sec\chi e^{-z})\}}} dz, \qquad \ldots\ldots(11)$$

where

$$f_0^2 = N_0 \epsilon^2/\pi m. \qquad \ldots\ldots(12)$$

The integral can be put in a simple form involving the single parameter

$$\frac{f_0^2 \cos^{\frac{1}{2}}\chi}{f^2} = \frac{f_c^2}{f^2}. \qquad \ldots\ldots(13)$$

It is convenient also to use the notation

$$b = \frac{f_c^2}{f^2} \sqrt{(2e)}, \qquad \ldots\ldots(14)$$

and to make the substitution

$$y = (\tfrac{1}{2} \sec\chi)^{\frac{1}{2}} e^{-\frac{1}{2}z} \qquad \ldots\ldots(15)$$

in (11). This then becomes

$$\int \kappa ds = -\frac{\nu_0 H}{\mathbf{c} \sec\chi} \int \frac{2by^2 e^{-y^2} dy}{\sqrt{(1 - bye^{-y^2})}}. \qquad \ldots\ldots(16)$$

The integral has to be evaluated numerically between limits specified by the practical problem under consideration. There are three important cases to be considered.

(i) *A wave for which $f<f_c$ is propagated vertically upwards from a level at which μ effectively unity*

In this case the wave is reflected from the level given by the greater root y_1 the equation

$$bye^{-y^2}=1. \qquad \ldots\ldots(17)$$

The value z_1 of z corresponding to this value of y is by (15)

$$z_1=\ln(\tfrac{1}{2}\sec\chi)-2\ln y_1, \qquad \ldots\ldots(18)$$

and the height h_1 of the level of reflection above the earth's surface is by (2)

$$h_1=h_0+H\ln(\tfrac{1}{2}\sec\chi)-2H\ln y_1. \qquad \ldots\ldots(19)$$

Values of y_1 for various values of f/f_c are given in table 1.

The total absorption in the wave thus reflected and returned to its startin point is by (16)

$$\frac{\nu_0 H}{c\sec\chi}F_1\left(\frac{f}{f_c}\right), \qquad \ldots\ldots(20)$$

where

$$F_1\left(\frac{f}{f_c}\right)=4b\int_{y_1}^{\infty}\frac{y^2e^{-y^2}\,dy}{\sqrt{(1-bye^{-y^2})}}. \qquad \ldots\ldots(21)$$

Numerical values of the function F_1 for various values of f/f_c are given table 1. Some remarks on the evaluation of the integral are made in § 7. Th integral is divergent for $f=f_c$.

Table 1

	Reflection from below			Reflection from above			
f/f_c	y_1	$F_1(f/f_c)$	$P_1(f/f_c)$	y_2	$F_2(f/f_c)$	$P_2(f/f_c)$	$p_1(f/f_c)$
0·0	∞	4·000	0·0	0·0	0·0	5·545	0·0
0·05	2·805	4·252	0·354	0·001	0·000	5·545	0·003
0·1	2·526	4·310	0·438	0·004	0·000	5·545	0·010
0·15	2·344	4·360	0·510	0·010	0·000	5·546	0·023
0·2	2·204	4·408	0·578	0·017	0·001	5·547	0·041
0·25	2·087	4·454	0·646	0·027	0·003	5·550	0·064
0·3	1·985	4·502	0·716	0·039	0·006	5·555	0·094
0·35	1·893	4·552	0·789	0·053	0·012	5·564	0·129
0·4	1·809	4·606	0·867	0·069	0·020	5·577	0·172
0·45	1·730	4·663	0·952	0·088	0·033	5·597	0·222
0·5	1·654	4·726	1·046	0·108	0·051	5·625	0·281
0·55	1·581	4·796	1·150	0·132	0·076	5·664	0·351
0·6	1·510	4·875	1·269	0·158	0·110	5·717	0·432
0·65	1·440	4·966	1·407	0·188	0·156	5·790	0·528
0·7	1·369	5·073	1·570	0·221	0·219	5·890	0·643
0·75	1·297	5·203	1·771	0·258	0·305	6·027	0·782
0·8	1·222	5·367	2·027	0·300	0·426	6·220	0·958
0·85	1·142	5·588	2·376	0·350	0·604	6·507	1·189
0·9	1·053	5·913	2·901	0·411	0·887	6·971	1·522
0·92	1·013	6·099	3·207	0·441	1·057	7·252	1·708
0·94	0·969	6·346	3·616	0·475	1·287	7·639	1·950
0·95	0·944	6·506	3·885	0·494	1·438	7·895	2·105
0·96	0·918	6·704	4·222	0·516	1·629	8·221	2·296
0·98	0·854	7·341	5·328	0·570	2·249	9·304	2·901

ii) *A wave for which $f > f_c$ is propagated vertically upwards from a level at which μ is effectively unity*

The wave will pass through the region and is assumed to be reflected downvards by a distinct upper region. The total absorption in the double passage of he lower region* is

$$\frac{\nu_0 H}{\mathbf{c}\sec\chi}F\left(\frac{f_c}{f}\right), \qquad \ldots\ldots(22)$$

vhere

$$F\left(\frac{f_c}{f}\right)=4b\int_0^\infty \frac{y^2e^{-y^2}\,dy}{\sqrt{(1-bye^{-y^2})}}. \qquad \ldots\ldots(23)$$

Numerical values of F are given in table 2.

Using (13) and (14), (22) may be put in the alternative form

$$\frac{\nu_0 H f_0^2}{\mathbf{c} f^2}(\sec\chi)^{-3/2}\,\Phi\left(\frac{f_c}{f}\right), \qquad \ldots\ldots(24)$$

vhere

$$\Phi\left(\frac{f_c}{f}\right)=\frac{\sqrt{(2e)}}{b}F\left(\frac{f_c}{f}\right). \qquad \ldots\ldots(25)$$

Values of the function Φ are also given in table 2. In the form (25) it appears

Table 2. Transmission

f_c/f	$F(f_c/f)$	$\Phi(f_c/f)$	$P(f_c/f)$	$p(f_c/f)$
0·0	0·000	4·133	0·000	0·000
0·05	0·010	4·136	0·010	0·003
0·1	0·041	4·146	0·042	0·013
0·15	0·094	4·164	0·094	0·030
0·2	0·168	4·188	0·169	0·055
0·25	0·264	4·221	0·267	0·087
0·3	0·384	4·262	0·390	0·127
0·35	0·528	4·313	0·540	0·177
0·4	0·700	4·374	0·720	0·236
0·45	0·900	4·447	0·934	0·308
0·5	1·134	4·534	1·187	0·394
0·55	1·403	4·639	1·485	0·497
0·6	1·715	4·765	1·839	0·621
0·65	2·078	4·917	2·261	0·771
0·7	2·501	5·105	2·771	0·956
0·75	3·005	5·341	3·398	1·189
0·8	3·616	5·649	4·195	1·493
0·85	4·387	6·073	5·256	1·911
0·9	5·436	6·711	6·796	2·543
0·92	5·992	7·080	7·658	2·909
0·94	6·689	7·570	8·780	3·396
0·95	7·120	7·889	9·496	3·713
0·96	7·637	8·287	10·375	4·108
0·98	9·183	9·562	13·125	5·378

---

* The absorption in the upper region is to be calculated from (20) and added to this.

that as $f_c/f \to 0$, $\Phi \to \sqrt{(2\pi e)} = 4{\cdot}133\ldots$, and thus if $f \gg f_c$, the total absorption

$$\frac{4{\cdot}13\nu_0 H}{\mathbf{c}} \frac{f_0^2}{f^2} (\sec\chi)^{-3/2}. \qquad \ldots\ldots(26)$$

This relation has been used by Best and Ratcliffe (1928) in a discussion of t experimental results. At lower frequencies it appears from (22) and (23) that t $(\sec\chi)^{-3/2}$ law in (26) does not hold, but it also follows quite generally that if f/f_c is constant, that is, if a frequency is chosen which is a constant multiple of the critic frequency, the total absorption varies as $\cos\chi$. Thus to study the variation absorption due to the variation in position of the sun it is desirable to experimen at a frequency which is a constant multiple of the critical frequency.

(iii) *A wave for which $f < f_c$ is propagated vertically downwards from above the lay from a level at which μ is effectively unity*

This case arises in an M reflection, or if waves are emitted from a source abov the layer. Since the Chapman region is unsymmetrical, the absorption in th case will be different from that for a wave of the same frequency reflected from th lower side of the region, and it is of some interest* to determine the amount this difference.

The level of reflection is now given by y_2, the least root of (17): values of thi are given in table 1.

The total absorption is given by

$$\frac{\nu_0 H}{\mathbf{c}\sec\chi} F_2\left(\frac{f}{f_c}\right), \qquad \ldots\ldots(27)$$

where

$$F_2\left(\frac{f}{f_c}\right) = 4b\int_0^{y_2} \frac{y^2 e^{-y^2}\,dy}{\sqrt{(1-bye^{-y^2})}}. \qquad \ldots\ldots(28)$$

Values of F_2 are given in table 1. It appears that the absorption is much les than that for a wave of the same frequency reflected from the lower side of the region.

§ 4. THE EQUIVALENT PATH

The equivalent path of the waves in the region is

$$\int \frac{ds}{\mu}, \qquad \ldots\ldots(29)$$

where μ is given by (5) and (7). Since, in all the cases in which we are intereste the path begins and ends in regions where μ is nearly unity, it is better to evaluat the excess of equivalent path over distance traversed, that is

$$\int\left(\frac{1}{\mu} - 1\right) ds. \qquad \ldots\ldots(30)$$

---

* It has been shown by Bagge (1943) that the ion density in the upper parts of the region will be reduced by diffusion to values much lower than the theoretical Chapman values. Thus the results given here will not correspond to the practical case at low frequencies.

Using the notation of § 3, and making the substitutions used there, this ecomes

$$-2H\int\frac{dy}{y}\left\{\frac{1}{\sqrt{(1-bye^{-y^2})}}-1\right\} \qquad \ldots\ldots(31)$$

Then in the three cases of § 3 the results are as follows:

ase (i). $f<f_c$. *Reflection from below*

The excess of equivalent path over distance traversed with which the wave turns to its starting-point is $HP_1(f/f_c)$, where

$$P_1\left(\frac{f}{f_c}\right)=4\int_{y_1}^{\infty}\frac{dy}{y}\left\{\frac{1}{\sqrt{(1-bye^{-y^2})}}-1\right\} \qquad \ldots\ldots(32)$$

d y_1 is the greater root of (17). y_1 and $P_1(f/f_c)$ are given in table 1. The actual uivalent path from the earth's surface to the layer and back is, by (19),

$$2h_0+2H\ln\left(\tfrac{1}{2}\sec\chi\right)-4H\ln y_1+HP_1(f/f_c). \qquad \ldots\ldots(33)$$

ase (ii). $f>f_c$. *Penetration*

The excess of equivalent path over distance in a double passage of the region $HP(f_c/f)$, where

$$P\left(\frac{f_c}{f}\right)=4\int_{0}^{\infty}\frac{dy}{y}\left\{\frac{1}{\sqrt{(1-bye^{-y^2})}}-1\right\}. \qquad \ldots\ldots(34)$$

Values of $P(f_c/f)$ are given in table 2.

ase (iii). $f<f_c$. *Reflection from above*

The excess of equivalent path over distance is $HP_2(f/f_c)$, where

$$P_2\left(\frac{f}{f_c}\right)=4\int_{0}^{y_2}\frac{dy}{y}\left\{\frac{1}{\sqrt{(1-bye^{-y^2})}}-1\right\}, \qquad \ldots\ldots(35)$$

d y_2 is the least root of (17). Values of y_2 and P_2 are given in table 1; it appears at P_2 is much greater than P_1.

. COMPARISON WITH RESULTS FOR A PARABOLIC DISTRIBUTION OF IONIZATION OVER A REGION OF THICKNESS $2y_m$

The parabolic distribution has been widely studied because of its simplicity. he results have been given by Appleton (1937) and are as follows.

In the case $f>f_c$ in which the wave penetrates the region, the excess of uivalent path over distance in a double passage is $y_m p(f_c/f)$, where

$$p(f_c/f)=\frac{2f}{f_c}\ln\frac{f+f_c}{f-f_c}-4. \qquad \ldots\ldots(36)$$

The quantity $p(f_c/f)$ is shown in the last column of table 2 for comparison with e corresponding quantity $P(f_c/f)$ for the Chapman region. Since equivalent ths in the two cases are expressed in terms of the thickness $2y_m$ and the scale eight H of their respective regions, direct comparison of the results is not possible, t the ratio $p:P$ may be seen to decrease steadily from 0·41 when $f_c/f=0{\cdot}98$ to 32 when $f_c/f=0{\cdot}1$.

It may be remarked that the Chapman distribution (5) has two points inflexion, and that the vertical distance between these is independent of χ ar equal to

$$H \ln \frac{2+\sqrt{3}}{2-\sqrt{3}} = 2{\cdot}63H.$$

Thus if a Chapman region were approximated to by a parabolic distribution, tl thickness $2y_m$ of the latter would be of the order of, but greater than, $2{\cdot}63H$.

For the parabolic distribution, if $f<f_c$, the wave is reflected at a height

$$y_m \frac{f_c-\sqrt{(f_c^2-f^2)}}{f_c} \qquad \text{......(37)}$$

above the bottom of the layer, and the equivalent path from the bottom of the lay to the level of reflection and back is

$$\frac{2y_m f}{f_c} \ln \frac{f_c+f}{f_c-f}. \qquad \text{......(38)}$$

From (37) and (38) the excess of equivalent path over distance for a wav reflected from the layer may be calculated. Writing this in the form $y_m p_1(f/f_c$ the quantity $p_1(f/f_c)$ is given in the last column of table 1 for comparison with th corresponding quantities $P_1(f/f_c)$ and $P_2(f/f_c)$ for reflection from the under an upper sides of a Chapman region. It appears that there are large differenc between the three curves.

§6. COMPARISON WITH EXPERIMENTAL P',f CURVES

There are two cases to be considered, according to whether the wave is reflecte from the lowest layer of the ionosphere or passes through this and is reflected fro a higher layer.

(i) *Reflection from a region of scale height H at height h_0*

In this case the equivalent path P' is by (33)

$$P' = 2h_0 + 2H \ln\left(\tfrac{1}{2}\sec\chi\right) - 4H \ln y_1 + HP_1(f/f_c), \qquad \text{......(39)}$$

where y_1 and P_1 are given in table 1. Measurements of an experimental P', curve at two values of f/f_c give two equations from which h_0 and H can be foun

(ii) *The wave passes through a region of scale height H_E and critical frequency f_E and is reflected from a higher region of scale height H at height h_0 whos critical frequency is f_c*

In this case

$$P' = 2h_0 + 2H \ln\left(\tfrac{1}{2}\sec\chi\right) - 4H \ln y_1 + HP_1(f/f_c) + H_E P(f_E/f), \qquad \text{......(40)}$$

where y_1 and P_1 are given in table 1 and P in table 2. In this case measuremen of three points on an experimental P', f curve will give the values of h_0, H and H_E

§7. EVALUATION OF THE INTEGRALS

This offers no particular difficulty. The integrals are readily transforme into forms suited to numerical integration, and series expansions are also availab which give satisfactory convergence for most of the values of f/f_c.

In the case of penetration the series expansions are

$$F(f_c/f)=\sum_{n=1}^{\infty} c_{n-1}b^n\,\Gamma(\tfrac{1}{2}n)n^{-\frac{1}{2}n}, \qquad \ldots\ldots(41)$$

$$P(f_c/f)=2\sum_{n=1}^{\infty} c_n b^n\,\Gamma(\tfrac{1}{2}n)n^{-\frac{1}{2}n}, \qquad \ldots\ldots(42)$$

here

$$c_n=\frac{1\cdot 3\ldots\ldots(2n-1)}{2\cdot 4\ldots\ldots(2n)}, \quad c_0=1. \qquad \ldots\ldots(43)$$

In the reflection case the roots y_1 and y_2 of (17) may be found from tables of $^{-y^2}$ or of the derivative of the error function. The series expansions are rather ore complicated; for example, for reflection from the lower side of the region,

$$P_1(f/f_c)=2\sum_{n=0}^{\infty}\frac{\Gamma(n+\frac{1}{2})c_{2n+1}b^{2n+1}}{(2n+1)^{(2n+1)/2}}\operatorname{erfc}[y_1(2n+1)^{\frac{1}{2}}]$$
$$+4\sum_{n=1}^{\infty}\frac{1}{(2y_1{}^2)^n}\sum_{r=2n}^{\infty}\frac{(r;\,n)c_r}{r^n}, \qquad \ldots\ldots(44)$$

here

$$(r;\,n)=(r-2)(r-4)\ldots\ldots(r-2n+2), \quad (r;\,1)=1,$$

nd c_n is defined in (43). The most important of the numerical series which ppear in (44) as coefficients of y_1^{-2n} can be summed accurately; the others are sily estimated numerically.

For reflection from the upper side of the region a series for y_2 is available romwich (1926), p. 161), and this leads easily to series expansions for F_2 and P_2.

§8. EXTENSION TO HIGHER VALUES OF THE COLLISION FREQUENCY

Throughout the above work it has been assumed that $\nu^2 \ll 4\pi^2 f^2$. This lation is normally satisfied in E- and F-regions, but it is probable that it will not old in D-region. This point is of some importance in connection with the study f absorption. Most workers, e.g. White and Straker (1939), use the Best and atcliffe formula (26) for the study of diurnal and seasonal variation of absorption, emarking that if much of the absorption takes place in D-region this formula, nd in particular the $(\cos\chi)^{3/2}$ law, will not hold good.

It is thus of some interest to study absorption at relatively high values of the ollision frequency. There is no difficulty in extending the above numerical ork to this case, but for a region in which μ is nearly unity, a simple extension of he Best and Ratcliffe formula (26) is easily obtained, and this should be sufficient or the discussion of experimental results.

To derive this we assume $f \gg f_c$, so that μ is nearly unity, but we do not neglect $^2/4\pi^2 f^2$. Then (8) is replaced by

$$\kappa=\frac{2\pi\epsilon^2 N\nu}{mc(4\pi^2 f^2+\nu^2)}.$$

Using the values (5) and (10) of N and ν, this gives for a double traverse of the egion

$$\int\kappa ds=\frac{4\pi\epsilon^2 N_0\nu_0 H}{mc}\int_{-\infty}^{\infty}\frac{\exp(\frac{1}{2}-\frac{3}{2}z-\frac{1}{2}\sec\chi e^{-z})\,dz}{4\pi^2 f^2+\nu_0{}^2 e^{-2z}}.$$

Making the substitution (15), this becomes

$$\int \kappa ds = \frac{H\nu_0 f_0^2}{cf^2}(\sec\chi)^{-3/2} I, \qquad \ldots\ldots(45)$$

where

$$I = 4\beta^2\sqrt{(2e)}\int_0^\infty \frac{y^2 e^{-y^2}\,dy}{y^4+\beta^2}, \qquad \ldots\ldots(46)$$

and

$$\beta = \frac{\pi f \sec\chi}{\nu_0}. \qquad \ldots\ldots(47)$$

It is found after some reduction that

$$I = 4\pi\beta^{\frac{3}{2}}e^{\frac{1}{2}}\{[\tfrac{1}{2} - C(\beta)]\cos\beta + [\tfrac{1}{2} - S(\beta)]\sin\beta\}, \qquad \ldots\ldots(48)$$

where $C(\beta)$ and $S(\beta)$ are the Fresnel integrals

$$C(\beta) = \frac{1}{\sqrt{(2\pi)}}\int_0^\beta \frac{\cos t}{\sqrt{t}}\,dt, \qquad S(\beta) = \frac{1}{\sqrt{(2\pi)}}\int_0^\beta \frac{\sin t}{\sqrt{t}}\,dt, \qquad \ldots\ldots(49)$$

which are tabulated in Jahnke-Emde (1933).

For large values of β, that is $2\pi f \gg \nu_0$, I has the value 4·133, and we regain (26). The behaviour of I for other values of β may be seen from table 3.

Table 3

β	I	β	I	β	I
0·0	0·000	2·5	3·093	8	3·931
0·5	1·109	3·0	3·287	10	3·996
1·0	1·919	4·0	3·547	15	4·067
1·5	2·463	5·0	3·706	20	4·095
2·0	2·828	6·0	3·810	∞	4·133

Since I is an increasing function of β and thus of $\sec\chi$, it follows that if $\int \kappa ds$ is plotted against $(\cos\chi)^{3/2}$ the resulting curve will not be the straight line given by (26) but will be concave downwards, the amount of the concavity being an indication of the amount of absorption taking place in regions where the collision frequency is high. The presence of a D-region of Chapman type will also have the effect of diminishing the ratio of summer to winter absorption.

Finally it should be remarked that although the effects of the earth's magnetic field have been neglected throughout, the results apply in the usual way to the important case of quasi-longitudinal propagation provided f is replaced by $f \pm f_L$.

ACKNOWLEDGMENT

I am indebted to Dr. F. W. G. White for introducing me to the subject, and for much helpful discussion; this work was begun in collaboration with him in the Division of Radiophysics of the Council for Scientific and Industrial Research, Australia.

REFERENCES

Appleton, E. V., 1937. *Proc. Roy. Soc.*, A, **162**, 451.
Bagge, E., 1943. *Phys. Z.*, **44**, 163.
Best, J. E. and Ratcliffe, J. A., 1928. *Proc. Phys. Soc.*, **50**, 233.
Bromwich, T. J. I'A., 1926. *Theory of Infinite Series*, Ed. 2 (Macmillan).
Chapman, S., 1931. *Proc. Roy. Soc.*, A, **132**, 353; 1931. *Proc. Phys. Soc.*, **43**, 26.
Jahnke-Emde, 1933. *Tables of Functions*, Ed. 2 (Teubner).
White, F. W. G. and Straker, T. W., 1939. *Proc. Phys. Soc.*, **51**, 865.

OBLIQUE RADIO TRANSMISSION IN THE IONOSPHERE, AND THE LORENTZ POLARISATION TERM

BY W. J. G. BEYNON,*
Communication from the National Physical Laboratory
* Now at University College, Swansea

MS. received 19 *September* 1946 †

ABSTRACT. The application of the Sellmeyer and Lorentz dispersion formulae to the problem of oblique reflection of radio waves from a parabolic type of layer is discussed. The work of Ratcliffe on the problem is extended and shown to be consistent with that of N. Smith. This theoretical analysis has been applied to a large number of experimental measurements of the maximum usable frequency at oblique incidence.

The results of two sets of measurements of the maximum usable frequency over distances of 1000 km. and 700 km. are in close agreement with values calculated from the Sellmeyer type of formula. In the case of the 1000 km. observations, the divergence between calculated and observed values is less than 0·2 of what it should be if the Lorentz formula applied. In the case of more accurate measurements over a distance of 700 km., the mean divergence between calculated and observed values is only about 0·03 of that which should occur on the basis of the Lorentz formula. These experimental results are thus in agreement with the theoretical conclusion of Darwin, that the Sellmeyer type of formula is applicable to the case of refraction of radio waves in the ionosphere.

---

§1. INTRODUCTION

THE question whether the Sellmeyer formula

$$\mu^2-1=\frac{-4\pi Ne^2}{mp^2}$$

or the Lorentz formula

$$\frac{3(\mu^2-1)}{\mu^2+2}=\frac{-4\pi Ne^2}{mp^2}$$

should be applied to the case of refraction in an ionised medium has been examined from the theoretical point of view by Darwin (1934, 1943). From the experimental side the problem has been considered by Booker (1938), Martyn (1938, 1939), Ratcliffe (1939), Smith (1941) and others. Darwin's theoretical analysis indicates that the simple Sellmeyer formula should be applicable, but the available experimental evidence has not yet been conclusive. From experiments on medium frequencies, Booker and Berkner concluded that it would be extremely difficult to interpret their results in terms of the Sellmeyer theory. Martyn and Munro also considered frequencies near the gyro-magnetic frequency, but concluded that the Sellmeyer theory was applicable. The experimental and theoretical work of Ratcliffe led to no conclusive result, whilst similar work by Smith again suggests that the Sellmeyer formula was appropriate to ionosphere theory.

In the present paper the theoretical work of Ratcliffe has been extended and shown to be consistent with that of Smith. The results of some oblique-incidence

---

† This paper was originally communicated to the Radio Research Board in 1943.

ionospheric measurements are also considered. These results appear capable explanation only in terms of the Sellmeyer formula.

§2. EARLIER WORK OF RATCLIFFE AND OF SMITH

The effect of including or omitting the so-called Lorentz polarisation term oblique incidence problems has been discussed by Ratcliffe (1939) and by Smit (1941). Ratcliffe considered an ionosphere with a parabolic gradient of ionisatio and concluded that at very short or very long distances there would be no marke difference between the magnitude of the maximum usable frequency calculate from either formula. He states that for region F_2 the greatest divergence betwee the two cases could only amount to about 2%. This would occur for a sende receiver distance of about 500 km., so that in practice the accuracy of measuremen would probably not be sufficient to decide between the two theories. In fac an attempt to apply the analysis to the results of some oblique-incidence exper ments for a sender-receiver distance of 464 km. was unsuccessful. On the oth hand, Smith (1941), from an analysis of the same problem, concluded that th divergence between the two calculated values of the maximum usable frequenc would progressively increase with increasing distance. For limiting distances o 1500 km. to 3000 km., Smith's analysis indicated that for transmissions b region E there should be a difference between maximum usable frequenci calculated from the Sellmeyer and Lorentz formulae of some 19% and for tran mission by region F_2 a difference of about 17%. From the calculations give below, it seems that the different results given by Ratcliffe and by Smith arose fro the fact that Ratcliffe confined his calculations to comparatively short distance Applied to larger distances the formulae of Ratcliffe show that the differen between maximum usable frequencies calculated from the Sellmeyer and Loren formulae is quite large and of the order stated by Smith.

§3. CALCULATION OF THE MAXIMUM USABLE FREQUENCY (M.U.F.) FACTOR CURVE

(a) *Case of plane earth and ionosphere. Lorentz formula*

It will be convenient to reconsider briefly the analysis given by Ratcliffe and t examine the extension of his formulae to greater distances. The effect of th earth's magnetic field will be neglected, and initially the earth and ionosphere ar considered to be plane. The nomenclature will be similar to that used b Ratcliffe. The ionosphere is assumed to have a parabolic gradient of ionisatio density, and the scale of height is arranged so that the semi-thickness is unity The Lorentz formula can then be written

$$\mu^2 = 2\left(\frac{z^2 - a^2}{b^2 - z^2}\right),$$

where μ is the refractive index at distance z below the level of maximum ionisatio density;

$$a^2 = 1 - \frac{f^2}{f^{02}}, \qquad b^2 = 1 + \frac{2f^2}{f^{02}};$$

f is the frequency under consideration and f^0 the normal-incidence critic frequency of the reflecting region.

The horizontal range D_1 corresponding to the portion of the path in the layer of a frequency f incident at angle i_0 (figure 1) is then given by

$$D_1 = 2\int_{\mu=1}^{\mu=s} \tan i \,.\, dz = 2\int_{z=1}^{z=c} \frac{\sin i_0 \,.\, dz}{\sqrt{\mu^2 - s^2}}$$

$$= 2s(2+s^2)^{-\frac{1}{2}}(1+2\epsilon^2)^{\frac{1}{2}}\Big[F(k,\chi) - E(k,\chi)\Big]_{\alpha}^{\pi/2}, \qquad \ldots\ldots(1)$$

where

$$s = \sin i_0, \qquad \epsilon = f/f^0,$$

$$\alpha = \sin^{-1}\sqrt{\frac{b^2-1}{b^2-c^2}} = \sin^{-1}\sqrt{\frac{2+s^2}{3}},$$

$$k^2 = \frac{6\epsilon^2}{(2+s^2)(1+2\epsilon^2)}, \qquad c^2 = \frac{2a^2+s^2b^2}{2+s^2}.$$

$F(k,\chi)$ and $E(k,\chi)$ can readily be evaluated from the usual tables of elliptic integrals.

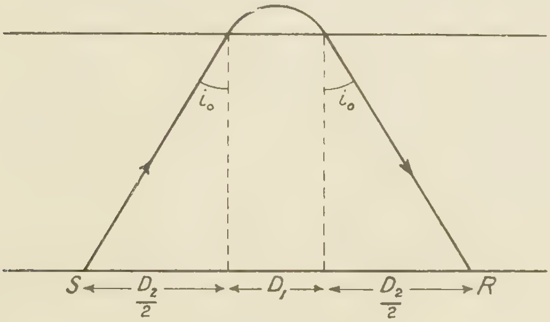


Figure 1.

The total range D is given by

$$D = D_1 + 2h_0 \tan i_0, \qquad \ldots\ldots(2)$$

h_0 being the height of the lower edge of the layer above the ground. We now plot the relation between D and i_0 for a particular fixed value of ϵ and so find when the range is a minimum. This gives us the maximum usable frequency for that particular range.

(b) *Case of plane earth and ionosphere. Sellmeyer formula*

For a parabolic layer in which the Lorentz term is assumed to be zero, the horizontal range is given by

$$D = D_1 + 2h_0 \tan i_0 = \epsilon \sin i_0 \log_e \frac{1+\epsilon\cos i_0}{1-\epsilon\cos i_0} + 2h_0 \tan i_0. \qquad \ldots\ldots(3)$$

In this case the M.U.F. curve can readily be deduced without plotting (D, i_0) curves by the method given by Appleton and Beynon (1940, 1947).

(c) *The equivalent height measured at normal incidence*

In practice the calculation of the M.U.F. factor curve is based on normal-incidence equivalent-height measurements. It is therefore necessary to examine the form of the (equivalent-height, frequency) curve to be expected from each of the two theories.

Lorentz formula

In this case Ratcliffe has shown that for a layer of parabolic type the equivalen height in the layer at normal incidence is given by

$$h'(\epsilon,0)=\sqrt{\frac{2}{1+2\epsilon^2}}\Big[(1+4\epsilon^2)E(k,\theta)-(1-2\epsilon^2)F(k,\theta)\Big]_{\theta=0\cdot956}^{\theta=\pi/2}-\tfrac{4}{3}k^2, \qquad \ldots\ldots(4)$$

k and ϵ having the same significance as before.

Sellmeyer formula

In this case the corresponding formula is

$$h'(\epsilon,0)=\frac{\epsilon}{2}\log_e\left(\frac{1+\epsilon}{1-\epsilon}\right). \qquad \ldots\ldots(5)$$

Figure 2 shows the equivalent height in the layer calculated from equations (4) and (5). It will be noted that the difference between the two cases is small. If we denote the semi-thickness of the region by y_m then the maximum difference at any frequency only amounts to about $0{\cdot}11y_m$. In fact the change from the Sellmeyer curve to the Lorentz curve can be simulated by small changes in the level of maximum ionisation density and in the semi-thickness of the region. Examination of figure 2 shows that if application of the Lorentz formula to any given experimental vertical incidence (h', f) curve, yields a level of maximum ionisation, say h_0+y_m and a semi-thickness y_m, then application of the Sellmeyer formula to the same curve would approximately yield a level of maximum ionisation of $h_0+0{\cdot}89y_m$

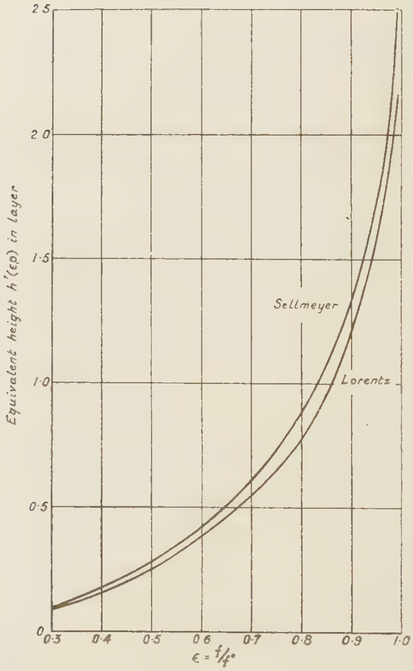


Figure 2. Normal incidence (h', ϵ) curves. Lorentz and Sellmeyer formulae.

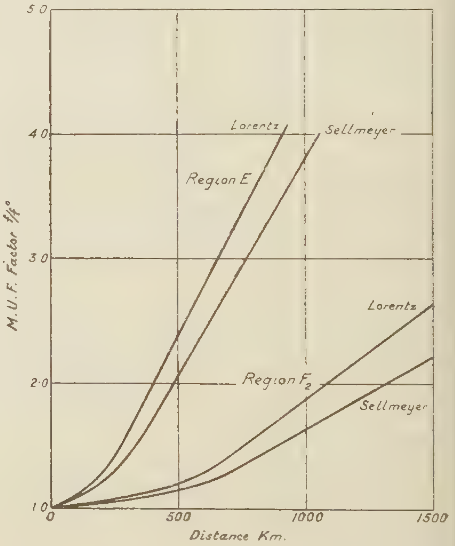


Figure 3. M.U.F. factor curves. Plane earth and ionosphere.

and a semi-thickness of $0{\cdot}93y_m$. Hence, in comparing M.U.F. factor curves calculated from the two theories, we must really compare the Lorentz M.U.F. factor curve for a level of maximum h_0+y_m and semi-thickness y_m with the Sellmeyer curve corresponding to a level of maximum $h_0+0{\cdot}89y_m$ and semi-thickness $0{\cdot}93y_m$. Figure 3 shows four M.U.F. factor curves, two for region E and two for typical region F_2 conditions. One curve in each pair corresponds to the Lorentz formula and one to the Sellmeyer formula. It will be noted that for distances less than about 500 km. for region F_2, the difference between the two cases is small, but the divergence steadily increases with distance, and at 1000 km. amounts to 12%. For region E the divergence is much larger than for region F at all distances. These calculations have assumed a plane earth and ionosphere, and rigid application of the results is really confined to distances for which earth-curvature effects are negligible. Any correction to these results for earth and ionosphere curvature will, however, affect the Sellmeyer and Lorentz curves to about the same order, and it is found that the ratios of M.U.F. factors to be deduced from the curves in figure 3 are really valid up to about 1000 km. We now consider some modifications of these formulae which will compensate approximately for the effect of earth and ionosphere curvature.

(d) *Case of curved earth and ionosphere. Lorentz formula*

Again we consider a parabolic type of layer, the height of the lower edge being h_0.

Let μ represent the refractive index at radial distance r from the centre of the earth. Then if χ represents the angle at the centre of the earth corresponding to the part range D_1 (figure 4) we have

$$D_1 = R\chi = R\int \frac{R_0 \sin i_0 \, dr}{r\sqrt{(\mu^2 r^2 - R_0^2 \sin^2 i_0)}}, \qquad \ldots\ldots(6)$$

$R_0 = R+h_0$, R being the radius of the earth.

As before we set

$$\mu^2 = 2\left[\frac{z^2-a^2}{b^2-z^2}\right].$$

If R_m be the radius of curvature corresponding to the level of maximum ionisation then $z=R_m-r$. Equation (6) then becomes

$$D_1 = R\int \frac{R_0 \sin i_0 \, dz}{(R_m-z)^2\left\{2\,\dfrac{z^2-a^2}{b^2-z^2} - \dfrac{R_0^2 \sin^2 i_0}{(R_m-z)^2}\right\}^{\frac{1}{2}}}. \qquad \ldots\ldots(7)$$

An explicit solution for D_1 necessitates some simplification of this integral, and in this connection it is relevant to note the magnitude of the horizontal range in which we shall be interested. In the present application of the formulae we shall only be concerned with transmission distances up to 1000 km. If we assume that the effect of earth and ionosphere curvature is of the same order of magnitude in the Sellmeyer and Lorentz cases, then it is easy to show that the curvature correction to the part range D_1 only amounts to about 15 km. for a total range of 1000 km. At the maximum range in which we shall be interested, the order of the curvature correction for the part range D_1 is thus only $1{\cdot}5\%$ of the total range. Hence, for present purposes a solution for D_1 which includes a first-order curvature correction will be adequate.

A solution of equation (7) valid to terms in $1/R_m$ is

$$D_1 = \frac{2R\,.\,R_0 \sin i_0}{R_m^2(2+s_0^2)^{\frac{1}{2}}}\left\{(1+2\epsilon^2)^{\frac{1}{2}}\Big[F(k,\chi)-E(k,\chi)\Big]_\alpha^\beta + \frac{3\sqrt{2}\,.\,s_0\epsilon^2}{[R_m c_0(2+s_0^2)^3]^{\frac{1}{2}}}\right\}, \qquad \ldots\ldots(8)$$

where $\alpha = \sin^{-1}\sqrt{\frac{2+s_0^2}{3}}$, $k^2 = \frac{6\epsilon^2}{(2+s_0^2)(1+2\epsilon^2)}$,

$\beta \simeq \frac{\pi}{2}$, $c_0^2 = 1 - \frac{2\epsilon^2(1-s_0^2)}{2+s_0^2}$,

$\epsilon = f/f^0$, $s_0 = \frac{R_0 \sin i_0}{R_m}$.

The derivation of this solution for D_1 is given in a note at the end of the paper

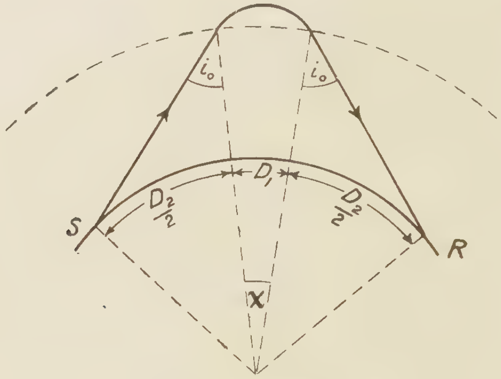


Figure 4.

The portion of the range corresponding to that part of the ray trajectory which is out of the ionosphere can be obtained from

$$D_2 = 2R\chi_2,$$

where χ_2 is given in

$$\chi_2 = \sin^{-1}\frac{[h_0 + R(1-\cos\chi_2)]\tan i_0}{R}. \qquad \ldots\ldots(9)$$

Equations (8) and (9) give the total range D, and maximum usable frequencies can now be calculated in the manner noted previously.

(e) *Case of curved earth and ionosphere. Sellmeyer formula*

In this case the solution for any type of parabolic layer has been given by Appleton and Beynon (1940, 1947), and details of the method have been described elsewhere.

Figure 5 shows M.U.F. factor curves for typical conditions for regions E and F_2, calculated from formulae (8) and (9) above. Corresponding curves for the Sellmeyer formula are also included in this figure. Table 1 gives the ratios of the M.U.F. factor calculated from the Lorentz formula to that calculated from the Sellmeyer formula for a series of distances and for reflections from region F_2. Column (2) in table 1 shows ratios calculated from the Lorentz and Sellmeyer formulae for a parabolic layer, with the constants $y_m = 100$ km. and $h_0 = 200$ km.

The values shown in column (3) compare M.U.F. factors for a layer with $y_m = 70$ km., $h_0 = 230$ km. in the Lorentz case, with M.U.F. factors for a layer with $y_m = 65$ km., $h_0 = 227$ km. in the Sellmeyer case (*vide* § 3 (c)). For purposes of comparison, ratios deduced from the work of Smith (1941) are also included in the table.

Table 1

Distance (km.)	Region F_2		
	$\dfrac{\text{M.U.F. factor (Lorentz formula)}}{\text{M.U.F. factor (Sellmeyer formula)}}$		Ratio of M.U.F. factors (Smith)
	$\left.\begin{array}{r} y_m = 100\text{ km.} \\ h_0 = 200\text{ km.} \end{array}\right\}$ (S & L)	$\left.\begin{array}{r} y_m = 70\text{ km.} \\ h_0 = 230\text{ km.} \end{array}\right\}$ (L) $\left.\begin{array}{r} y_m = 65\text{ km.} \\ h_0 = 227\text{ km.} \end{array}\right\}$ (S)	
500	1·01	1·01	1·01
700	1·06	1·06	1·04
1000	1·13	1·13	1·12
1500	1·19	1·17	1·15
2000	1·20	1·18	1·16

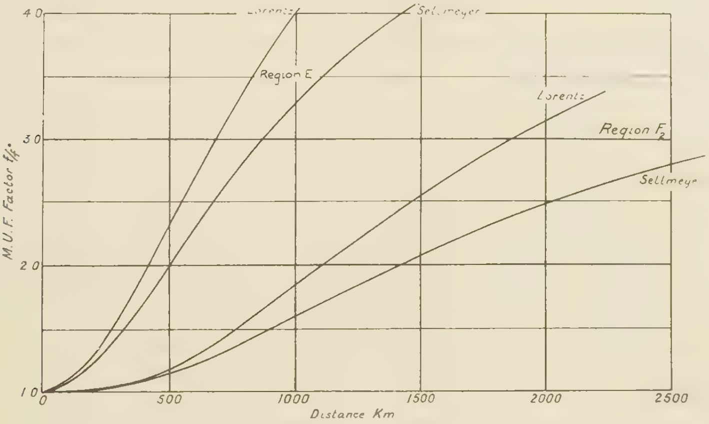


Figure 5. M.U.F. factor curves. Curved earth and ionosphere.

For any given reflecting region the magnitude of these ratios depends to a small extent on the particular values which are assumed for the layer parameters y_m and h_0, but it will be noted that the present calculations are in substantial agreement with those of Smith. At 700 km. we should expect the M.U.F. factor calculated from the Lorentz formula to exceed that calculated from the Sellmeyer formula by about 15% for region E and by 4% to 6% for region F_2. At 1000 km. the difference amounts to 17% for region E and to 13% for region F_2. The divergence between the results of Ratcliffe and of Smith appears to have arisen because Ratcliffe's calculations were confined to the shorter distances at which the differences between values calculated from the two dispersion formulae for region F_2 are certainly very small.

§4. COMPARISON WITH EXPERIMENTAL RESULTS

Reference to the curves of figure 5 shows that the divergence between the tw theories is most marked in the M.U.F. factor curves for region E. Accurat measurements of the M.U.F. of region E should, therefore, provide a more sensitiv test for deciding between the two theories than similar measurements on region F_2 In practice, however, measurements of the M.U.F. of the normal region E ar seldom possible because, in this latitude at least, oblique-incidence reflection from normal region E are often masked by abnormal region-E reflections. Th extremely variable nature of abnormal region-E reflections both at normal and a oblique incidence makes any accurate measure of the M.U.F. factor for this regio extremely difficult, if not impossible. The experimental results are therefor only concerned with measurement of the maximum usable frequency of region F_2

In two recent series of experiments, details of which are in course of publication, measurements of the maximum usable frequency for region F_2 have been made over distances of 1000 km. and 700 km. respectively, and we now consider the results of these experiments in relation to the foregoing analysis.

(a) *Observations on Zeesen broadcasting senders*

In these experiments observations were made on the signal received at Slough from the Zeesen high-frequency broadcasting station situated approximately 1000 km due east of the receiving site. Vertical-incidence ionospheric data were available only at the receiving end of the transmission path, and conditions at the mid-point were deduced from these observations. Values of the maximum usable frequency were estimated from the times of appearance and disappearance of the ray signal at sunrise and sunset. From a group of thirty-two observations it was concluded that the mean calculated value of the M.U.F. was less than that observed by 3·2%. It was realised, however, that the accuracy of these measurements was limited because of the lack of accurate ionospheric data at the mid-point of the trajectory, and much of the discrepancy seemed due to this cause. From the results of a group of twenty-one observations which were considered most reliable, it was deduced that the mean difference between observed and calculated M.U.F.s did not exceed about 2%. Now from table 1 it will be noted that for a sender-receiver distance of 1000 km. we should expect a difference of 12 to 13% between maximum usable frequencies calculated from the Lorentz and Sellmeyer formulae respectively. The fact that, in actual practice, agreement at least to within about 2% was obtained with values calculated on the Sellmeyer formula, indicates that this is the formula which is applicable to the ionosphere.*

(b) *Actual measurement of the maximum usable frequency using pulse technique*

In these experiments a large number of accurate observations of the maximum usable frequency over a sender-receiver distance of 715 km. were made using pulse technique. Vertical-incidence data obtained simultaneously with the oblique observations were available at both ends of the transmission path. The frequency radiated by the pulse sender could be varied at will and maximum usable frequency measurements made at any time. The use of pulse technique nabled the maximum usable frequency of either the ordinary or extraordinary omponent to be actually observed visually or photographically recorded, and this, ogether with the fact that adequate ionospheric data for both ends were available, ives these results considerably greater accuracy than those quoted above.

---

* The results of a subsequent series of similar measurements over a distance of 2500 km. are also consistent with this conclusion.

Here again the observed maximum usable frequency was compared with that alculated from the vertical incidence ionospheric data on the basis of the Sellmeyer ormula. The calculated value was taken to be the arithmetic mean of that calulated from the prevailing ionospheric conditions at the two ends of the path.

For a series of 110 measurements it was found that 50 observations showed an verage error of $+3{\cdot}8\%$, 56 observations an average error of $-3{\cdot}75\%$ and 4 bservations showed negligible error. Occasionally the divergence was much in xcess of these average figures, but positive and negative discrepancies were equally requent, and it is fairly certain that these errors were due to inaccurate knowledge f conditions at the mid-point of the trajectory. From the results of these 110 xperiments it was found that the mean of all the calculated values agreed with the mean observed value to within about $-0{\cdot}2\%$. For this distance of transmission he Sellmeyer and Lorentz formulae yield calculated M.U.F.s differing by about 6% able 1). The divergence from the Sellmeyer calculated value is thus approximately only 1/30th of what it should be if the Lorentz formula is applicable. The xtremely small divergence between experimental values and the values calculated n the basis of a Sellmeyer formula again agrees with the conclusion that this is the ype of formula which is valid for the ionosphere.

It may be noted that in both series of experiments the calculated value is lightly less than that observed, but it is doubtful whether this difference is significantly different from zero. Further experimental measurements will be necessary efore this point can be settled. In any case the divergence is less than one-fifth f what is required if the Lorentz formula is applicable.

§ 5. CONCLUSIONS

The results of two series of measurements on the maximum usable frequency ver distances of 1000 km. and 700 km. are in very close agreement with values alculated in the basis of the Sellmeyer dispersion formula. The results are in greement with those given earlier by Smith and with the theoretical conclusions f Darwin.

APPENDIX

Note on calculation of part-range D_1 (Lorentz formula. Curved ionosphere)

We have

$$D_1 = R\chi = \frac{2R\,R_0 \sin i_0}{R_m^2}\int_{z=z_1}^{z=1} dz\left(1-\frac{z}{R_m}\right)^{2}\left\{2\,.\,\frac{z^2-a^2}{b^2-z^2}-\frac{R_0^2\sin^2 i_0}{R_m^2}\left(1-\frac{z}{R_m}\right)^{-2}\right\}^{-\frac{1}{2}}. \qquad \ldots\ldots(10)$$

To terms in $1/R_m$ this can be written

$$D_1 = \frac{2R\,R_0\sin i_0}{(2+s_0^2)^{\frac{1}{2}}R_m^2}\left\{\int_{z_1}^{1}\left(\frac{b^2-z^2}{z^2-c_0^2}\right)^{\frac{1}{2}}dz+\frac{1}{R_m}\left[\int_{z_1}^{1}\left(\frac{b^2-z^2}{z^2-c_0^2}\right)^{\frac{1}{2}}.\,dz^2\right.\right.$$

$$\left.\left.+\frac{s_0^2}{2(2+s_0^2)}\int_{z_1}^{1}\left(\frac{b^2-z^2}{z^2-c_0^2}\right)^{3/2}.\,dz^2\right]\right\}, \qquad \ldots\ldots(11)$$

where

$$s_0 = \frac{R_0 \sin i_0}{R_m}, \qquad c_0^2 = \frac{b^2 s_0^2 + 2a^2}{2+s_0^2},$$

and z_1 is given by

$$\frac{z_1^2 - c_0^2}{b^2 - z_1^2} = \frac{2 s_0^2 z_1}{R_m(2+s_0^2)}.$$

The three integrals in equation (11) can readily be evaluated. An expression fo D_1 accurate to terms in $1/R_m$ is then found to be

$$D_1 = \frac{2R\,R_0 \sin i_0}{R_m^2[2+s_0^2]^{\frac{1}{2}}}\left\{(1+2\epsilon^2)^{\frac{1}{2}}\Big[F(k,\chi)-E(k,\chi)\Big]_\alpha^\beta + \frac{6\epsilon^2}{R_m(2+s_0^2)}\left[\frac{\theta}{2}\left(\frac{4-s_0^2}{2+s_0^2}\right) + \frac{\sin 2\theta}{4}\left(\frac{4+s_0^2}{2+s_0^2}\right) - \frac{s_0^2}{2+s_0^2}\cot\theta\right]_{\theta_1}^{\theta_2}\right\}, \quad \ldots\ldots(12)$$

where

$$\alpha = \sin^{-1}\sqrt{\frac{2+s_0^2}{3}},$$

$$\beta = \sin^{-1}\sqrt{\frac{(1+2\epsilon^2 - z_1^2)(2+s_0^2)}{6\epsilon^2}} \simeq \frac{\pi}{2},$$

$$\theta_1 = \sin^{-1}\sqrt{\frac{z_1^2 - c_0^2}{b^2 - c_0^2}} \simeq \sin^{-1}\sqrt{\frac{2s_0^2\, c_0}{R_m(2+s_0^2)}},$$

$$\theta_2 = \sin^{-1}\sqrt{\frac{1-s_0^2}{3}}.$$

$$k^2 = \frac{6\epsilon^2}{(2+s_0^2)(1+2\epsilon^2)}.$$

It will be noted that θ_1 is very small and the principal term in $1/R_m$ in equation (12 is that containing $\cot\theta_1$.

Now

$$\cot\theta_1 = \sqrt{\frac{b^2 - z_1^2}{z_1^2 - c_0^2}} = \sqrt{\frac{R_m(2+s_0^2)}{2s_0^2\, z_1}} \simeq \sqrt{\frac{R_m(2+s_0^2)}{2s_0^2\, c_0}}.$$

Substitution in (12) thus gives

$$D_1 = \frac{2R\,R_0 \sin i_0}{R_m^2(2+s_0^2)^{\frac{1}{2}}}\left\{(1+2\epsilon^2)^{\frac{1}{2}}\Big[F(k,\chi)-E(k,\chi)\Big]_\alpha^\beta + 3s_0\epsilon^2\sqrt{\frac{2}{R_m c_0(2+s_0^2)^3}}\right\}. \quad \ldots\ldots(13)$$

An examination of the magnitude of the term of order $1/R_m$ in either equation (12 or (13) shows that the error in D_1 introduced by neglecting this term is very small. It is found that for a layer of semi-thickness 100 km., terms in $1/R_m$ in equation (12 contribute 2 km. to a total range of 700 km., 5 km. to a total range of 1000 km. and 18 km. to a total range of 2000 km. For a thinner layer the contribution of these terms is proportionately less. It is thus to be expected that values of the total range involving the part-range D_1 calculated from equation (13), should be accurat to better than 0·1% at all distances. Furthermore, the error involved in neglecting rms of order $1/R_m$ and using the simple formula

$$D_1 = \frac{2R\, R_0 \sin i_0}{R_m^2 (2+s_0^2)^{\frac{1}{2}}} (1+2\epsilon^2)^{\frac{1}{2}} \Big[F(k,\chi) - E(k,\chi) \Big]_{\alpha}^{\beta} \qquad \ldots\ldots(14)$$

approximately only 0·5% for a total distance of about 1000 km.

ACKNOWLEDGMENT

The work described in this paper was carried out as part of the programme of e Radio Research Board and the results are published by permission of the epartment of Scientific and Industrial Research.

REFERENCES

PPLETON, E. V. and BEYNON, W. J. G., 1940. "The Application of Ionospheric Data to Radio-Communication Problems."—Part I. *Proc. Phys. Soc.*, **52,** 518; Part II, 1947. *Ibid.*, **59,** 58.

OOKER, H. G. and BERKNER, L. V., 1938. "Constitution of the Ionosphere and the Lorentz Polarisation Correction." *Nature, Lond.*, **141,** 562.

ARWIN, C. G., 1934. "The Refractive Index of an Ionised Medium." *Proc. Roy. Soc.*, **146,** 17; Part II, 1943. *Ibid.*, **182,** 152.

ARTYN, D. F. and MUNRO, G. H., 1938. "The Lorentz Polarisation Term and the Earth's Magnetic Field in the Ionosphere." *Nature, Lond.*, **141,** 159; 1939. "The Lorentz Polarisation Correction in the Ionosphere." *Terr. Mag. Atmos. Elect.*, **44,** 1.

ATCLIFFE, J. A., 1939. "The Effect of the Lorentz Polarisation Term in Ionospheric Calculations." *Proc. Phys. Soc.*, **51,** 747.

MITH, N., 1941. "Oblique Incidence Transmission and the Lorentz Polarisation Term." *Bur. Stand. J. Res. Wash.*, **26,** 105.

ON THE SPECTRA OF CS AND CSe

BY H. G. HOWELL,
Technical College, Bradford

MS. received 22 *February* 1945; *in revised form* 9 *September* 1946

BSTRACT. The accepted view that the main CS and CSe band systems are $^1\Pi - {}^1\Sigma$ nalogous to the Fourth Positive bands is criticized, and an alternative interpretation is put orward that they are $^3\Pi - {}^1\Sigma$ transitions, possibly the counterpart of the CO Cameron ands with the $^3\Pi$ level possibly perturbed by a $^1\Pi$ level.

§1. INTRODUCTION

THE majority of diatomic band systems involve the ground state of the molecule, and even in cases where the nature of the lower state is uncertain (as with certain oxide emission spectra) it is probably the ground state. Consequently, when the spectrum contains a number of band systems, it is possible to acquire more extensive and accurate data for this state than for the excited states. In comparatively few instances are excited states definitely involved in more than one system, and so our knowledge of them is meagre and fragmentary compared with our well-ordered catalogue of atomic states.

It is possible to systematize the data by correlating the states of one molecu with the corresponding ones of an analogous molecule, and such a correlation shou immediately reveal any missing levels for which a search can then be mad This process is simplified if rotational analyses have shown the nature of th states concerned; but if these are not available, more empirical methods must b used as a basis for the correlation. This note deals with such a method, and, as a result, the correctness of the present accepted interpretation of the spectra of CS and CSe is questioned.

§2. CORRELATION OF EXCITED STATES USING VIBRATIONAL FREQUENCIES

An attempt to correlate all states of the analogous molecules PbO, PbS SnO and SnS was made by Howell (1936) by comparing the correspondin vibrational frequencies. He showed that although the ratio ω''/ω' (which wil now be called Ω) varies irregularly from one excited state to the next for a given molecule, the same value of this ratio occurs for both oxide and sulphide molecul of the same element (Pb or Sn). This numerical agreement between states was taken to indicate a probable correspondence of state type, and it seemed to provide a reasonable method of correlation in the absence of rotational analyses. This comparison did not include the lighter molecules of this Group, viz. CO, CS, SiO, etc., as it was felt that close correspondence between the molecules of Ge, Sn and Pb (which are known to constitute a closely related group) was more likely. However, the extension to these molecules was made by Barrow and Jevons (1938) in their study of the spectrum of the group molecule SiS, and their table 7, with Ω instead of $1/\Omega$, is reproduced here as table 1, together with additional data from the later papers of Barrow (1939, 1940) and Barrow and Jevons (1940). This later work has brought in the selenides and tellurides, and it will be seen that the correspondence in Ω is wider than Howell first announced and seems sufficiently good with the exception of CS and CSe to justify the belief that the states involved are indeed corresponding ones. Barrow and Jevons consider that this correspondence does not imply that the states are all necessarily of the same theoretical type such as all $^1\Sigma$ or all $^1\Pi$ because, as they point out, whereas certain of the levels concerned are $^1\Pi$ others are accepted as being $^1\Sigma$. If this is true then there does not seem any point in making this comparison of Ω values and the correspondence found must be fortuitous and meaningless. If all the levels involved (i) have the same electron configuration, (ii) have the same type of coupling, and (iii) are free from strong perturbations, then a close correspondence between ω values can be anticipated, for condition (i) being fulfilled with these analogous molecules, the excitation of the electron to the same type of level will be accompanied by corresponding changes in the force constants from orbital to orbital, changes which will be reflected in the ω values. As for condition (ii), change of coupling is possible and it may be that the coupling of the C compounds is different from that of the Pb molecules which will probably exhibit case-c tendencies, but no change can be expected within the C and Si inner group nor in the Sn–Pb group. Consequently there is a possibility that the values of these two sub-groups refer to different levels, and the numerical agreement here found is accidental, but this point is not important for the present argument. Neglecting m (iii) for the moment, the writer believes that the agreement among the Ω lues is significant and does indicate similarity of term type (possibly modified a coupling change). Consequently any discrepancy such as occurs with both S and CSe indicates that corresponding levels are not concerned, and also that cepted views on term type which conflict with this opinion should be carefully rutinized. The cases of CS and CSe form the subject of this note, and the olecules PbO etc., having apparently all $^1\Sigma$ states, will be dealt with in another per.

§3. ELECTRON LEVELS OF CS AND CSe

In table 1 the only serious deviations from the average Ω value of 1·45 are those CS and CSe, which are 1·20 and 1·25 respectively. The CS system has been cepted as $^1\Pi - {}^1\Sigma$ (analogous to the Fourth Positive system of CO) chiefly as a sult of the rotational analysis by Crawford and Shurcliff (1934). Examination their paper immediately reveals certain weaknesses in their interpretation and ovides additional stimulus to finding an alternative :

Table 1. Ω values of Group IV oxides, etc.

	O	S	Se	Te
C	1·43	1·20	1·25	—
Si	1·46	1·46	1·44	1·43
Ge	1·51	1·53	1·49	1·47
Sn	1·41	1·47	1·52	1·49
Pb	1·45	1·52	1·50	1·46

(1) The writers themselves are uneasy about the number of disturbing omalies in their presumed $^1\Pi - {}^1\Sigma$ systems, e.g. under high dispersion “practi- lly every band is found to possess double *R*-branch heads, displaced origins, normal $\nu_{\mathrm{head}} - \nu_{\mathrm{origin}}$ values and irregular branches.” Two sets of *P*, *Q* d *R* branches are also noted.

(2) A scrutiny of their published spectrogram of the 0,1 band reveals other anches apparently overlooked by them.

(3) In addition to the main system, another briefer one with $\nu_e \sim 40000\ \mathrm{cm}^{-1}$. analysed as $^1\Sigma - {}^1\Sigma$ with the lower-state ω identical with ω_e'', thus making it nost certain that the ground state is involved. Yet the B'' value of this state ffers by 25% from that obtained from the main system.

They interpret the main system as a $^1\Pi - {}^1\Sigma$, with the upper state perturbed a $^3\Pi$ state. If this is so, then under condition (iii) it would be possible to count for the discordant Ω values. However, the large extent of the anomalous atures spread over practically all the bands indicates such a violent perturbation at the writer has sought for any possible alternative explanation. A $^3\Pi - {}^1\Sigma$ ansition is immediately suggested, in which case the system would be analogous the Cameron bands of CO. This would possibly account for all the observed mplexities and certainly for the increased number of branches. Indeed, mparison of the structure of the Cameron bands as analysed by Gerö, Herzberg d Schmid (1937) with that of CS reveals quite a close resemblance. Also the per $^3\Pi$ state of CO has an Ω value of 1·25 against the 1·20 of CS. The main objection to this view is one of intensity—such a transition should be much weak than the corresponding $^1\Pi - {}^1\Sigma$, but in this connection the experiments of Knau and Cotton (1931) on the intensities of the CO bands are very illuminatin They observed that whereas high-pressure excitation conditions favoured th production of the Fourth Positive, low-pressure conditions tended to suppre them whilst increasing the intensity of the Cameron system.

This selective effect of pressure on the singlet and triplet systems can b expected to occur also in CS and CSe, with the triplet system becoming relative stronger in the direction CO—CS—CSe owing to the increase in molecul weight. Consequently, such a $^3\Pi - {}^1\Sigma$ transition should have a higher probabili with CS and CSe than with CO if the conditions are suitable. Now CS ban can be obtained under high- or low-pressure conditions, and Martin (191 observed that when they were produced in a carbon arc the most intense ban were around 2579 A., and these were suppressed in the Geissler discharge whic is used for the normal CS system. Crawford and Shurcliff actually state tha very low pressures were necessary in their work in order to eliminate the extensiv S_2 spectrum, and also that higher pressures decreased their intensity. It seem therefore, that the conditions eminently suit the production of a $^3\Pi - {}^1\Sigma$ system

The intensity objection would then appear to have lost much of its strength It can also be pointed out that the extent of the so-called perturbation of the $^1\Pi$ means a strong interaction of the $^3\Pi$ and $^1\Pi$ which is violating the singlet-triple prohibition just as much as in a $^3\Pi - {}^1\Sigma$ transition and that any objection to th intensity of the $^3\Pi - {}^1\Sigma$ system applies equally well to this perturbation.

The presence of the 2579-A. bands (which appear to have been forgotten b subsequent workers) would indicate the close presence of another state and possibl perturbation. Consequently the correct interpretation of the so-called $^1\Pi - {}^1\Sigma$ system should be among the following:—(*a*) a $^3\Pi - {}^1\Sigma$ without perturbation (*b*) $^3\Pi - {}^1\Sigma$ with the upper state perturbed by $^1\Pi$; (*c*) $^1\Pi - {}^1\Sigma$ with the uppe state perturbed by $^3\Pi$.

If (*b*), the upper state of the 2579-A. bands can be identified as the $^1\Pi$. Fro its Ω value, the only known excited state of CSe is analogous to that of CS jus discussed, but no rotational analysis is available. Examination of Barrow' paper (1940) on the vibrational analysis shows that in attempting to interpret th system as $^1\Pi - {}^1\Sigma$, the same difficulties are encountered as with CS. For example it is necessary to introduce a cross term in $u'u''$ to account for the systemati deviations of ΔG values, due apparently to the variation in interval between R head and origin which should surely be small for this molecule as well as for CS Even with this additional term it is also found necessary to postulate a vibrationa perturbation for $v'=1$, and finally a few bands cannot be definitely allocated. It would seem that a transition of the type $^3\Pi - {}^1\Sigma$ would account for certain o these difficulties.

§4. CONCLUSION

It would appear that there is room for more experimental investigation o these molecules, particularly of the 2579-A. bands of CS. Also, even if th alteration in the upper state type of these molecules, suggested here, is incorrect, he difference in the B'' values found by Crawford and Shurcliff still remains to be accounted for.

REFERENCES

Barrow, R. F., 1939. *Proc. Phys. Soc.*, **51**, 269.
Barrow, R. F., 1940. *Proc. Phys. Soc.*, **52**, 380.
Barrow, R. F. and Jevons, W., 1938. *Proc. Roy. Soc.*, A, **169**, 45.
Barrow, R. F. and Jevons, W., 1940. *Proc. Phys. Soc.*, **52**, 534.
Crawford, F. H. and Shurcliff, W. A., 1934. *Phys. Rev.*, **45**, 860.
Gerö, L., Herzberg, G. and Schmid, R., 1937. *Phys. Rev.*, **52**, 467.
Howell, H. G., 1936. *Proc. Roy. Soc.*, A, **153**, 683.
Knauss, H. P. and Cotton, J. C., 1931. *Phys. Rev.*, **38.**, 1190.
Martin, L. C., 1913. *Proc. Roy. Soc.*, A, **89**, 127.

A SIMPLE OPTICAL MODEL DEMONSTRATING THE PRINCIPLE OF THE BRAGG X-RAY SPECTROMETER

By F. A. B. WARD,
The Science Museum, London S.W. 7

MS. received 22 *February* 1946; *demonstrated* 5 *July* 1946

ABSTRACT. The model illustrates by an optical analogy the fact that a monochromatic beam of X rays will be reflected by a single crystal only when the crystal is set at a particular angle to the incident beam such that reflections from successive sheets rich in atoms are in phase and so give an interference maximum. The direction of the reflected beam can also be determined.

§ 1. GENERAL DESCRIPTION OF THE MODEL

The model is designed for exhibition in the Science Museum to illustrate the principle of the Bragg X-ray spectrometer. It can be constructed, however, of such simple materials and at such a low cost that it should be of general use for teaching the fundamental principle of X-ray analysis of crystals.

As shown in the photograph, the incident beam of X rays is represented by a sheet of corrugated iron, 18 × 12 in. The spacing between successive crests is 3 in., so that six complete waves can be seen. It is advantageous, though not essential, to paint the system black and to paint a white strip $\frac{1}{4}$ in. wide along the crest of each wave.

The crystal to be analysed is represented by four glass plates (12 × 10 in. photographic negatives with the emulsion removed, the longer dimension being vertical). The plates are mounted one behind the other, 2 in. apart, in a simple wooden rack; an additional grooved distance piece is placed on top of the plates to ensure more accurate parallelism. The rack is provided with a plywood hood in order to cut out stray light. The rack, whose horizontal dimensions are $10\frac{3}{4}$ × 7 in., is pivoted on a central spigot, and carries two pointers indicating upon a scale of degrees.*

---

* In the Science Museum model this consists of a 12-in. celluloid protractor placed over a white ivorine disc.

In use, the corrugated iron is illuminated from above by means of an adjustable table lamp, and the observer looks into the glass plates. Images of the corrugated iron are seen by reflection from each plate, but in general the crests of the corrugations will not be in phase, and a confused picture is seen. If, however, the plate-rack is rotated until the relation $n\lambda = 2d \sin\theta$ is satisfied, where λ is the distance between corrugations and d is the spacing of the plates, then all the reflected images of the crests will be in phase and the images will coalesce, giving a simple image as of a single corrugated sheet. θ is then the angle at which a reflected x-ray beam would be obtained.

The principle is thus illustrated, but for more accurate measurements it is advisable to replace the corrugated iron by a flat wooden board of similar dimensions, black but with white lines $\frac{1}{4}$ in. wide, spaced at 3-in. intervals. These do not so obviously represent waves, but they can be more accurately spaced than the crests of the corrugated iron, which depart slightly from a true sine-wave form. As before, the plate-rack is rotated until the successive white lines appear accurately superposed. In order to evaluate θ, readings α_1 and α_2 are taken on either side of the incident beam; the angle $\alpha_2 - \alpha_1$, through which the plate-rack has been turned, is then equal to $180° - 2\theta$, so that

$$\theta = 90° - \frac{\alpha_2 - \alpha_1}{2}.$$

Having determined θ, one can either assume λ to be known and evaluate d or vice versa.

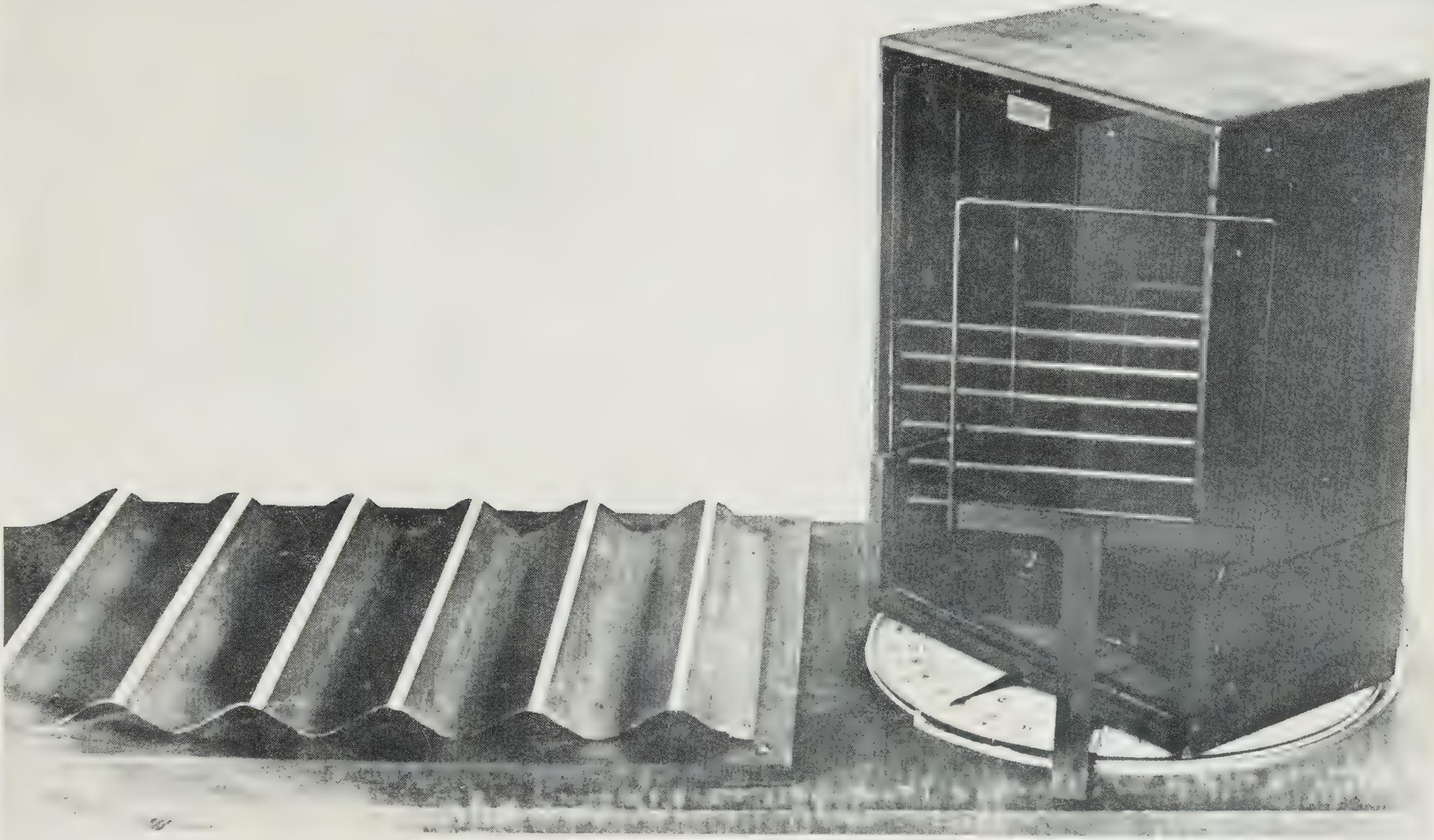
To facilitate the interchange of the corrugated iron and painted wood, the former is mounted on a wooden board, in each corner of which a $\frac{1}{2}$-in. hole is bored. These holes fit over wooden dowel pins projecting from the general baseboard. Similar holes are bored in the corners of the blackened wooden board carrying the white lines.

In order to find the direction of the reflected beam, the model is provided with an extra arm, as shown in the photograph. This arm rotates about the same spigot as the plate rack; it is bent upwards and carries a horizontal straight-edge, perpendicular to the axis of rotation, and 6 in. above the baseboard. In use, this arm is rotated until the straight-edge is seen to be parallel to the images of the wave-crests; the edge is then perpendicular to the reflected beam, whose direction can be read off by means of a pointer mounted on the arm and indicating on the circular scale. It will be seen that a simple wire view-finder is mounted above the straight-edge.

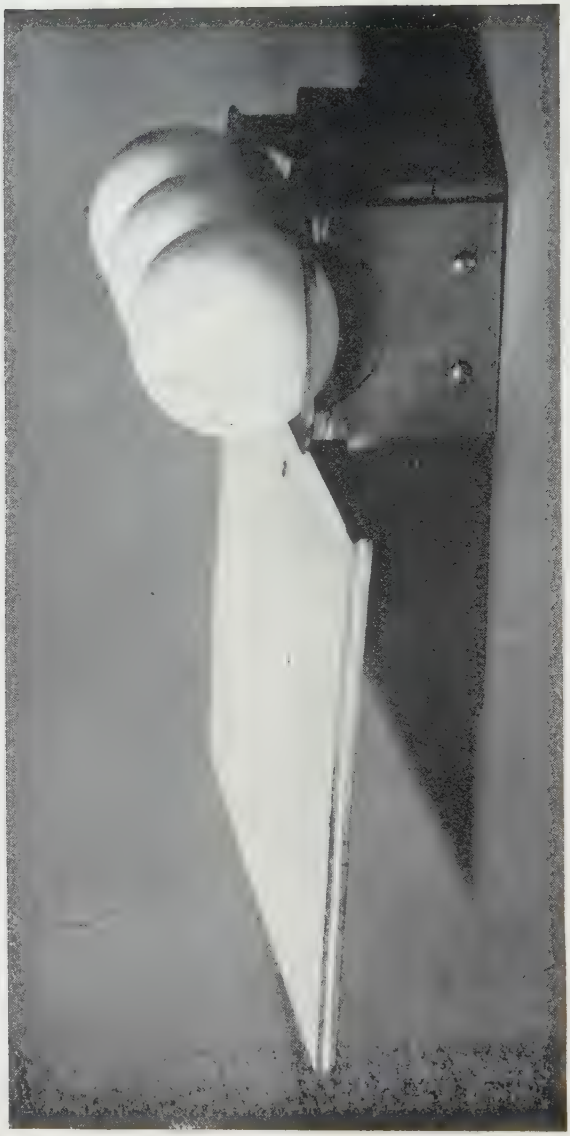
§2. LIMITATIONS OF THE MODEL

The finite thickness of the glass plates causes a doubling of each image, but, provided that the thickness of each plate is a small fraction of the spacing of the plates, this does not interfere seriously with the efficiency of the model; it was for this reason, however, that it was made on such a relatively large scale.

It would appear possible to design a smaller model, using corrugated cardboard instead of corrugated iron, and glass plates of quarter-plate size, one side of each plate being "coated" to reduce reflection in the manner now employed in various optical instruments and camera lenses. The coating would, of course,



A simple optical model demonstrating the principle of the Bragg x-ray spectrometer.



A single unit of the uranium chain reaction model.

produce its maximum effect only at a particular angle of incidence, and it would be preferable to illuminate the incident “wave” with monochromatic or approximately monochromatic light.

ACKNOWLEDGMENT

I wish to thank Dr. H. Shaw, Director of the Science Museum, for permission to publish this paper.

A MECHANICAL MODEL ILLUSTRATING THE URANIUM CHAIN REACTION

By F. A. B. WARD,
The Science Museum, London S.W. 7

Demonstrated 5 *July* 1946; *MS. received* 16 *October* 1946

ABSTRACT. The model consists of 30 units, each representing the nucleus of a uranium atom. Each unit consists of two portions, held together by a form of catch which, when released, allows the two portions to fly apart, while at the same time three table-tennis balls, representing neutrons, are projected upwards to a height of two to three feet.

To demonstrate the model, the 30 units are placed in rows in a glass showcase measuring × 3 × 3 ft. high. A single table-tennis ball is dropped upon the trigger of one of the units; this unit immediately disintegrates, projecting its three “neutrons”, which impinge upon neighbouring units, causing them to break up and initiating a “chain reaction” which spreads until a majority of the units have disintegrated, the whole process occupying a few seconds.

The units are designed and their layout planned so as to secure maximum efficiency.

§ 1. GENERAL DESCRIPTION

THE model was designed for demonstration in the Atomic Energy Exhibition which opened at the Science Museum on 14 February 1946. It is intended to illustrate the typical uranium chain reaction which has been utilized for the release of atomic energy. The model consists of a number of units, each representing the nucleus of an atom of ^{235}U or of plutonium. In designing the units, experiments were first tried with a breakback mouse-trap. These showed promising results, and from them the type of unit now employed was evolved. The detailed design of this unit, an example of which is shown in the photograph, is due to Mr. G. H. C. Jones, Deputy Foreman of the Metal-Working Shop at the Science Museum.

Each unit consists of two portions, shown in figures 1 and 2 respectively, which can be coupled together in a manner to be described. The first of these portions, shown in plan and elevation in figure 1, consists of two blocks of wood, A and B, hinged together at C, and fitted with short projecting steel pins D and E. The blocks are normally kept apart by the helical springs F and G, their separation being limited by the head of the retaining screw Z. To couple the two portions of a unit together, ready for operation, the upper block A is pressed down by hand, and the two pins D and E are inserted into the holes H and I in the other portion of the

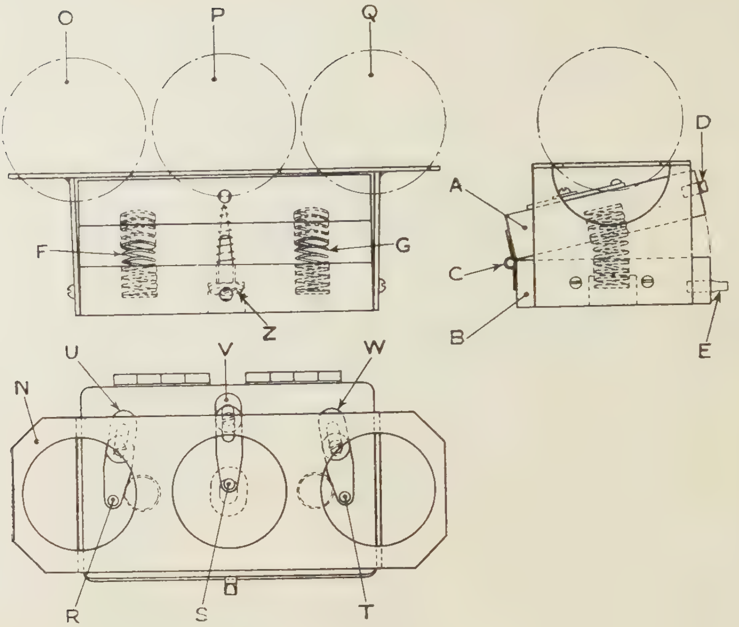


Figure 1.

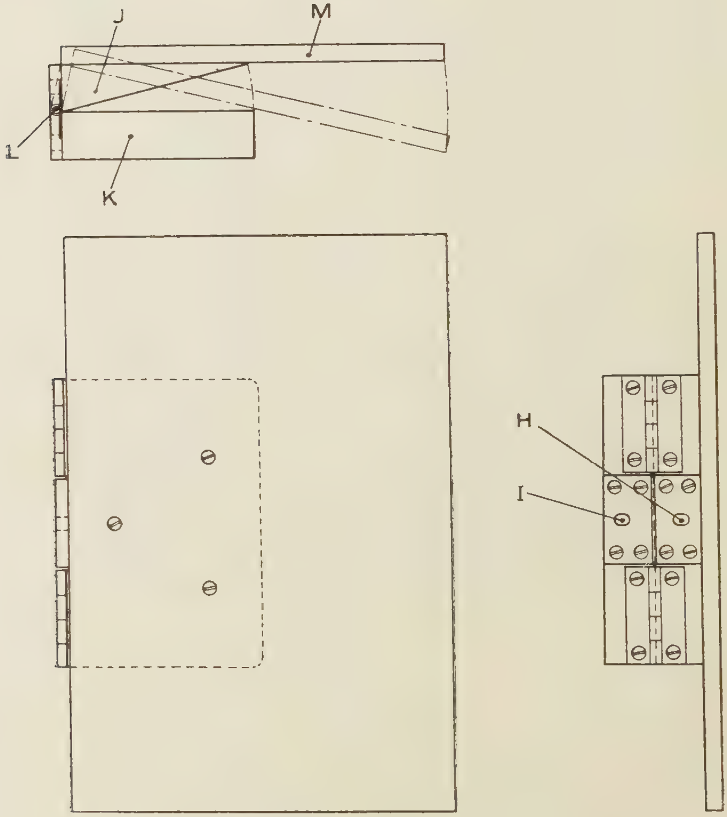


Figure 2.

it. This second portion consists of two wooden blocks, J and K, hinged at L. rectangular "target" M, of thin plywood, is screwed to the upper block J and, shown in the photograph, lies approximately in a horizontal plane when the it is "set".

When a table-tennis ball falls upon this target, it pushes it down from the sition shown by the full lines in figure 2 into that shown by the chain-dotted lines, d releases the upper pin D from the hole H. The helical springs F and G then ce the upper block A of the first portion of the unit smartly upwards, until it is pped by the head of the retaining screw Z. As it rises, it strikes the three le-tennis balls O, P and Q, which rest upon circular holes in a brass frame N ich is screwed to block B by means of side flanges. In the unit as first designed, balls were struck directly by the upper surface of A but, in order to give some ntrol over the flight of the balls, three adjustable strikers were later fitted. ese consist of small steel rivet-heads R, S and T mounted on adjustable slotted ss strips U, V and W. In the absence of the strikers, the balls are projected an angle of about 20–30° from the vertical, with speed sufficient to carry them to a height of four to five feet, so that they fall several feet away from the unit. e strikers can, however, be adjusted so as to project the balls more nearly tically by means of a more glancing blow which also gives a lower speed of jection. In practice, the strikers are arranged so that the balls are projected to a height of two to three feet, and fall about six to twelve inches away from the t.

In setting up a unit, it is found best to push the projecting pins of the trigger y home into their sockets and then to ease them a little, so as to make the trigger re sensitive.

To demonstrate the chain reaction, the units are lined up in rows on the base- rd of a standard museum showcase whose interior dimensions are 6 × 3 × 3 ft. h. The lay-out adopted is shown in figure 3.

The first row contains a single unit, the second row four units, and the remain- rows five each. In the first four rows, all units are arranged so as to throw the ls forward to the next row; but in later rows some units are arranged to throw kwards, in order to set off any earlier units which may have escaped being struck. the units in the last row are arranged to throw backwards, for obvious reasons.

The reaction is initiated by dropping a single table-tennis ball upon the unit in first row. The releasing mechanism is extremely simple, and is roughly strated in figure 4, which shows pictorially the general course of the reaction. e ball to be dropped is placed inside a short cylindrical brass tube, and is mally prevented from falling by a transverse diametral pin passing through es in the sides of the cylinder. The pin can be withdrawn by pulling a string ched to one of its ends. The ball then falls upon the target of the unit in the t row; this unit explodes, projecting its three "neutrons" upwards and to the ht (see figure 4); some of these "neutrons" impinge upon the targets of other ts, which also blow up, and the "chain reaction" is established. If a relatively e number of hits happen to be scored in the first few rows, the action becomes ost "explosive" in type, and lasts only about three seconds, but, if the early nber of hits is smaller, the release of energy proceeds more steadily, lasting 10 seconds.

§2. NUMERICAL REQUIREMENTS FOR THE MAINTENANCE OF THE "CHAIN REACTION"

It is of interest to determine the conditions limiting the development of th chain reaction.

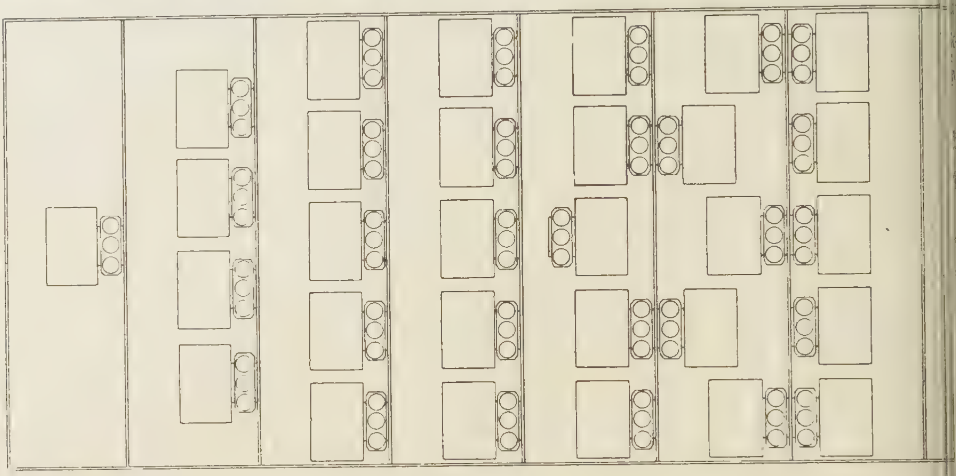


Figure 3.

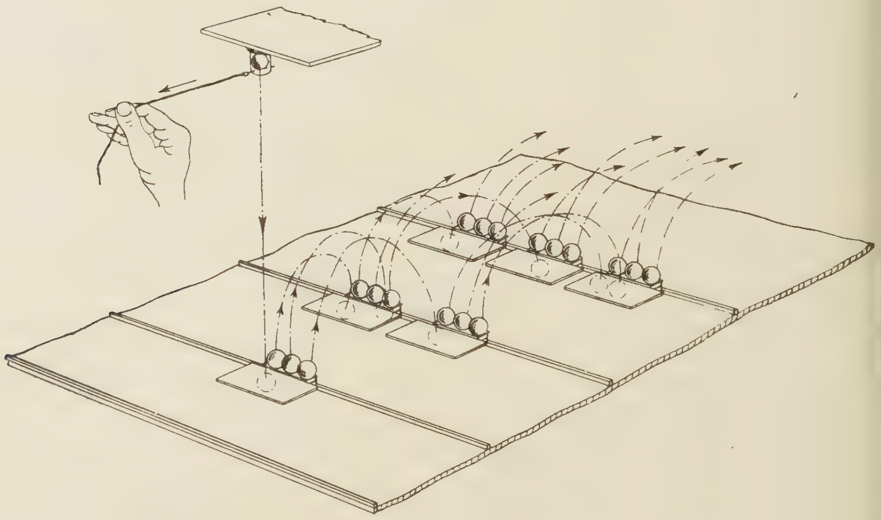


Figure 4.

The target area of each unit measures 6 × 4 in. and, if the unit is proper adjusted, any ball striking this area will produce a disintegration. If the ba strikes any other part of a unit—such as the table-tennis balls or their supporti frame—it makes an inelastic collision and is lost to the reaction. Each unit h accordingly what we may term a " dead area ", which measures a little less th 6 × 4 in.

In a disintegration, the two halves of a unit fly apart two to three inches befo coming to rest, and hence the units cannot be placed too close together, or they w set each other off by direct collision, and the analogy with the uranium reaction w spoiled. With the layout shown in figure 3, the total floor space occupied is –80 sq. in. per unit. Of the space occupied by the model, therefore, roughly e-third represents live target area, one-third “ dead ” area of the units them- lves, and one-third area not covered by units. It can easily be seen that, if is last area is non-reflecting to table-tennis balls, a chain reaction will be difficult maintain. For consider the fate of the three balls projected from the first it disintegrated. The chance that *one* of these will score a miss is 2/3. The ance that all three will miss is $(2/3)^3$ or 8/27. In 8 cases out of 27, therefore, can expect that the reaction will not even start.

If, however, the floor space not occupied by units is totally reflecting, so that the lls rebound from it to their original height, the chance that one of the balls ojected from the first unit will score a miss is 1/2, for it is equally probable that ien the ball strikes a unit it will strike the “ live ” or the “ dead ” part of it. In is case, therefore, the chance of all three balls missing, and the reaction failing start, is reduced to 1 in 8, with a corresponding increase in the probability that, ce started, it will develop.

In the actual model, the baseboard of the case containing the model is of painted od, from which the balls rebound to a height of about two-thirds of that from ich they fall. This is sufficiently close to total reflection to enable the reaction proceed satisfactorily. If the ball is effective after one bounce, but not after o, it can be shown that the probability of all three balls from the first unit missing, d the reaction failing to start, is $(5/9)^3$ or approximately 1/6. In practice, the sults are more favourable than either the figure 1/6 or 1/8 would suggest, and a mplete failure is very rare indeed. This is mainly because the strikers of the st unit can be so adjusted that the balls have a very good chance of falling on the rgets of the second row—they can, in fact, be definitely “ aimed ” at these targets; is also probable that the “ dead ” area of each unit has been over-estimated in the ove analysis, and furthermore is not entirely “ dead ”.

In a single demonstration, the average number of units disintegrated is about out of the total of 30. As the reaction proceeds, the floor of the case becomes tered with disintegrated units, so that the proportion of “ dead ” area begins to crease rapidly, targets of disintegrated units being now “ dead ” area. One nnot therefore normally expect to disintegrate all 30 units, though it can be done relatively rare occasions when chance is fortunate.

ACKNOWLEDGMENTS

I wish to express my thanks to Dr. H. Shaw, Director of the Science Museum, r permission to publish this paper. I am much indebted to Mr. G. H. C. Jones, ho was responsible for the detailed design of the units. I also wish to thank r. S. H. Groom for valuable criticisms and suggestions relating to the lay-out the units.

THE EMISSIVITY OF HOT METALS IN THE INFRA-RED

By DEREK J. PRICE,
S.W. Essex Technical College, London

Communicated by H. Lowery; *MS. received* 15 *July* 1946

ABSTRACT. A specially designed vacuum-tight, water-jacketed furnace has been use in the measurement of the emissivity of incandescent metals for wave-lengths of 1·0 to 4·5 Results have been given for pure platinum, iron, molybdenum, copper and nickel. A " spl cylinder " type of comparison black-body has been used and its efficiency has been compare with a spherical standard. This calibration has suggested a correction to be applied Ives' standard of visible radiation. From the present emissivity values, and from thos given by other workers, it appears that for a wave-length (usually in the near infra-re peculiar to each metal the emissivity is constant over a large range of temperature. Suc *X-points* have been observed in the case of iron and molybdenum, but not with platinum o copper.

§ 1. INTRODUCTION

The study of radiation emitted from hot objects is important for at leas two reasons. On the one hand, such a study is fundamental to the practic of optical and total radiation pyrometry, and on the other, the developmen of modern physics has made it clear that the theory of radiation is a basic sectio of our knowledge of atomic and electronic mechanism. The " perfect black body " of which the properties are described by Planck's law of radiation doe not correspond with any real substance : indeed, a classical example has bee cited by Coblentz (1908).

> " It is, of course, absurd to attempt to measure the temperatures of these substance (complex minerals) with an optical pyrometer. For example, the rod of oligoclase was perfectly transparent glass and emitted no light on heating to over 1200° c., except that du to the sparking of the platinum terminals. Nevertheless, such substances as iron oxide a the same temperature would have emitted visible light, while both emit strongly in th infra-red region."

A number of attempts have been made to formulate directly the radiatio laws of imperfect radiators but these have not met with much success. It i more profitable to consider the degree of imperfection of the radiator, takin as a numerical measure the ratio of the energy radiated by an imperfect an a perfect body respectively under the same conditions. This ratio, havin a value between 0 and 1, defines the *emissivity* of the material. Kirchhoff' law (stating that the sum of the emissivity and reflectivity of an opaque bod must be unity) acts as a connecting link between the theory of optical constants

Hagen and Rubens (1900) at the end of the nineteenth century stimulate interest in this field by providing and testing successfully a theory of emissivit valid for long wave-lengths. Their equation $E=2\sqrt{\nu/\sigma}$ also accounted for th observed variation of emissivity with temperature since the change of σ (th electrical conductivity) with temperature was known independently for th metal used; ν is the frequency of the radiation. For wave-lengths of the order of 10μ or more, good agreement with this theory was found, but for shorter wave-lengths it failed completely. The temperature variation, too, being derived from the Hagen–Rubens relation, broke down for shorter wave-lengths. More recent work has shown that in the visible region some metals at least appear to have a temperature coefficient opposite in direction to that predicted by the simple theory. For all metals, the magnitude of this coefficient is very much smaller for short wave-lengths than that predicted by the Hagen–Rubens relation. For those few metals for which sufficient data have been obtained, it seems that there is in the near infra-red a wave-length for which the temperature coefficient is zero over a very large range of temperatures. This remarkable occurrence of a special wave-length (here called the *X-point*) peculiar to each of these metals does not appear to have been examined critically by any previous worker, although the possible existence of such a point has been noted by Worthing (1926). On the basis of two observed values he suggested that the X-point wave-length was proportional to the melting-point of the metal. This is at variance with other known X-points. The present work has been concentrated on the provision of more accurate data for the transition region between the range of validity of the Hagen–Rubens relation and the X-point in hope of further elucidating this phenomenon.

§ 2. PREVIOUS METHODS

The main methods that have been used to measure the emissivity of metals may be divided into five categories as follows :—

(1) Evaluation from measurement of the optical constants.
(2) Determination of reflectivity.
(3) Measurement of the true and apparent temperatures of a hot substance.
(4) Measurement of the quantity of energy radiated by a hot body at a known temperature.
(5) Direct determination as the ratio of the energies radiated by the hot body and a comparison black-body under the same conditions.

The first method is theoretically superior to the others since it yields two independent quantities, n and k, of which the emissivity is only a function. Against this, however, is the difficulty of catoptric observations in the infra-red and the added complication of working at temperatures as high as was desired. The second method has been more widely used than any other on account of its simplicity. There are three objections to its use in the present case. First, the hot metal mirror will radiate energy which must be compensated for by a method similar to that used by Beekman and Oudt (1925). Secondly, at high temperatures, the development of crystal facets and the warping of the mirror both tend to interfere with specular reflection. Thirdly, as has been pointed out by Hurst, a 1% error in the measured reflectivity may, in the infra-red, be equivalent to a 20% error in the calculated emissivity. This is due of course to the high reflectivity in this region of the spectrum. The measurement of true and apparent temperature is a very convenient method for use in the visible spectrum but difficulties become almost insuperable in the infra-red. The fourth method was abandoned because of the difficulty of making temperature and absolute energy readings of a sufficiently high order of accuracy. Preliminary calculations (Price and Lowery, 1944) have shown that the results obtained for instance by Hase using this method could not be satisfactorily converted to emissivity values.

The last method is free from all of the previously noted faults and presents only one difficulty, namely the provision of a comparison black-body. This has been the subject of some ingenuity in previous work, at least four different arrangements having been used to ensure equality of temperature between the specimen and black-body. Historically, Mendenhall (1911) was the first to use a method of this nature. His specimen was in the form of an electrically heated strip of thin metal bent to the shape of a V of small angle. Due to multiple reflection, black-body conditions obtain on sighting into the opening of the V. An improvement of this was made by Hurst (1933) who used a block of metal with a V cavity cut in it. This obviated the buckling of the metal which, in Mendenhall's case, proved to be the greatest source of inaccuracy. The disadvantage of Hurst's method is that for each specimen a specially machined block must be made. Other types of comparison black-body that have been used include a hole in a tubular filament (Langmuir, 1916) and a helical filament or strip (Worthing, 1925). Neither of these devices yields a conveniently shaped source for spectroscopic observation. Another device having all the advantages of Hurst's wedge apparatus, but allowing an easy change of specimen, was used by Drecq and later investigated by Ives (1924) as a primary standard of luminous radiation. These workers used a thin sheet of platinum bent into the form of a cylinder with the edges left slightly apart so as to form an axially placed slit. The cylinder was heated ohmically by means of copper electrodes at either end, to which it was rigidly clamped. It was decided to investigate whether this device could be readily adapted to working in an inert atmosphere or a vacuum in order to observe oxidizable metals.

§ 3. EXPERIMENTAL ARRANGEMENT

The first model furnace was made in the form of a small brass box enclosing a specimen 5 cm. long and 1 cm. diameter, heated by a current of about 300 amp. obtained from a single turn secondary on a 1 kVA. mains transformer, and observed through a small window of fused silica. The performance of this trial model may be summarized as follows :—

(1) Satisfactory results may be obtained with fixed terminal blocks since the deformation produced by thermal expansion of the specimen is small.
(2) The slight deformation so produced does not noticeably affect the radiation from the black-body slit.
(3) The centre portion of the specimen is at a reasonably uniform temperature.
(4) It is advisable to work *in vacuo* or in a stationary atmosphere since the specimen temperature is extremely sensitive to draughts.
(5) The temperature of the furnace becomes quite high (300° c.) after running for a short while and this destroys the efficiency of most forms of vacuum sealing cement.

A second model furnace of more massive design was then constructed with the vacuum joints formed by surface-ground steel plates bolted together. Some preliminary results were obtained with this apparatus, but the rapid heating up of the furnace body caused so many difficulties that a third model was constructed incorporating water cooling for the furnace, the electrodes, and the transformer secondary turn. It was made in two halves : a steel base plate and a one-piece cover of fused silica incorporating a window of optical quality. The design can be seen from figure 1. The electrodes were sealed in with vacuum wax and the joint between the two halves of the furnace was kept smeared with vacuum grease. The furnace was mounted on a board attached to the transformer and the whole assembly formed a rigid unit. This apparatus was used without alteration for all subsequent work. Its completely satisfactory performance may be summarized as follows :—

(1) The water cooling was so efficient that no stray heat could be detected and ordinary paraffin wax could be used for temporary seals.

(2) A good vacuum was obtained even with specimen temperatures of 1500° c.

(3) A cylindrical specimen could be made from sheet metal and inserted in the furnace ready for emissivity readings to be taken within ten minutes.

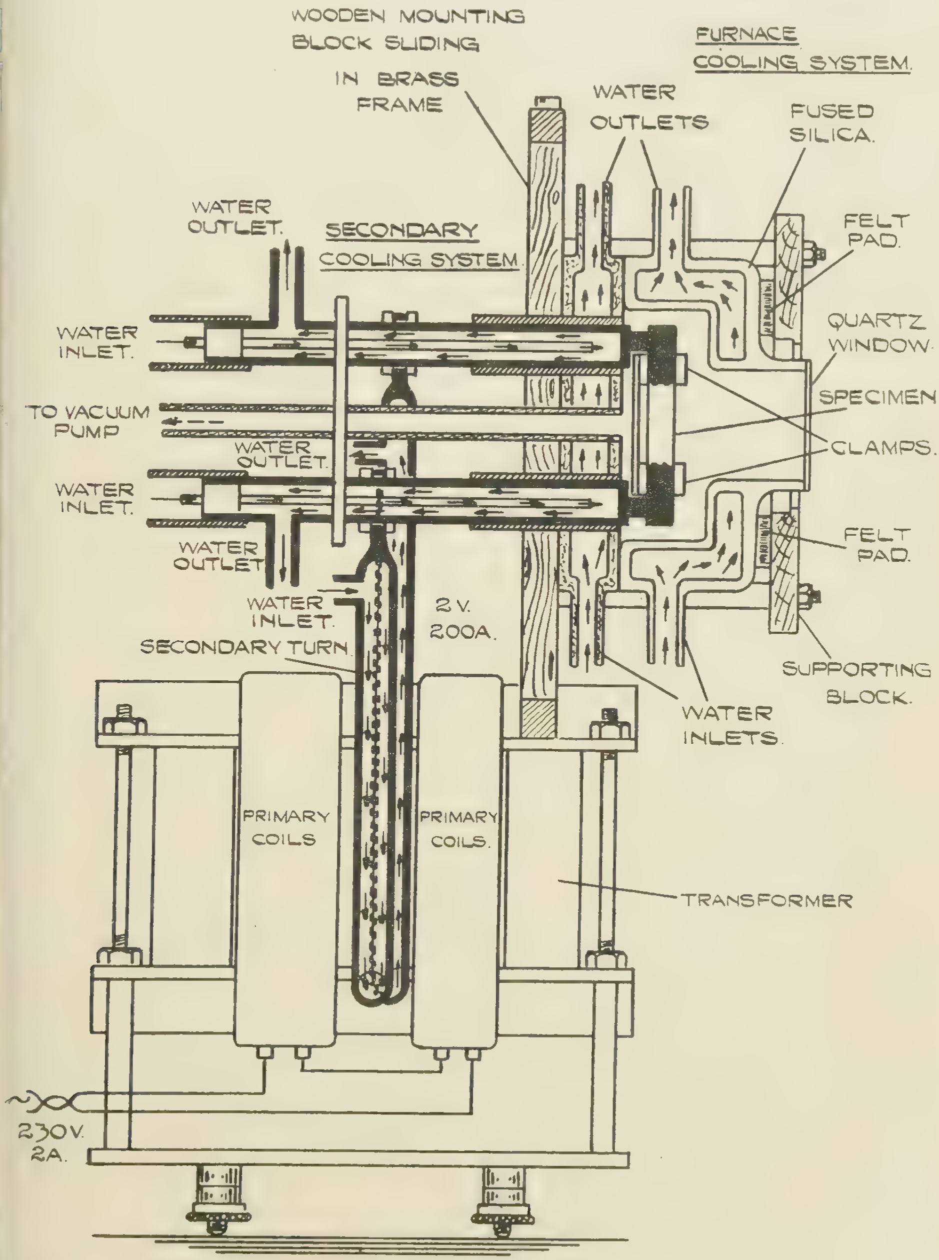


Figure 1. Emissivity furnace with auxiliary equipment.

This permitted the use of a large number of specimens in order to average out stray variations in the surfaces, etc.

A Schwartz-type thermopile mounted in air was used in conjunction with a Hilger Barfit Infra-red Spectrometer employing a rock-salt prism in a constant-deviation mounting. The thermopile was connected through a series resistance box to a Tinsley 4500 LS galvanometer. The complete arrangement was free from serious trouble both during the initial calibrations and throughout the subsequent experiments. The spectrometer and furnace, together with a large mirror for focusing the radiation from the furnace on the slit of the instrument, were mounted in a draught-tight case containing trays of calcium chloride to safeguard the rock-salt prism. A system of gearing and an observation window enabled the wave-length drum of the spectrometer to be adjusted from outside. Another window at the end of the case enabled an optical pyrometer to be sighted on the specimen inside the furnace. The focusing mirror was fitted with a slow-motion screw to rotate it so as to set either the black-body slit or the specimen surface adjacent to it on to the spectrometer slit. A series resistance and multi-tapped transformer enabled the input to the furnace transformer and therefore the temperature of the specimen to be controlled externally. A shutter controlled by a cable lead was fitted in front of the spectrometer slit. The vacuum tube from the furnace was connected by means of a two-way tap to a Hivac vacuum pump and to a cylinder of pure hydrogen (99·8 % H_2) supplied by the British Oxygen Co. By this means the furnace could be evacuated, flushed or filled with hydrogen. In later experiments an auto-transformer mains stabilizer was used to obviate fluctuations of the heating current that occurred during long series of observations.

§4. EXPERIMENTAL TECHNIQUE

The raw material for the specimens was in each case a sheet of metal about 0·2 mm. thick. After selecting a uniform area, a rectangle 5 cm. × 3 cm. was cut with a clean razor blade. This was then filed to the exact size needed for the required slit width, using a steel former. The sheet was then bent into a cylinder between two brass formers round a rod and subsequently annealed. After polishing with very fine emery (when this was necessary) the specimen was washed in alcohol, dried, and kept over calcium chloride until required. The polishing process was not often used since it was found that after a specimen had been used once at a high temperature, and removed from the furnace, the specimen surface had a very good, polished appearance due no doubt to local evaporation taking place on the scratches. It was found, however, that this "self-polishing effect" produced no measurable change in emissivity, and hence the effect of small surface scratches was disregarded. Prior to the fitting of each specimen, the electrodes and clamps were cleaned with emery cloth to ensure good contact. The specimen having been fitted, the furnace was evacuated for a few minutes to remove moisture and then the furnace was filled with hydrogen and evacuated again. After repeating this process and filling again with hydrogen, the furnace was ready for use. The furnace current was switched on, adjusted to the required value, and pyrometer readings were then taken on the specimen surface and on the black-body slit. At this stage any large buckling of the specimen could be observed and corrected by opening the furnace. This, however, was not a frequent cause of trouble. After this, emissivity readings were taken up and down the wave-length range at intervals of 0·1, 0·25 or 0·5 μ. Finally the pyrometer was read again as before. Each emissivity reading, with necessary repeats, involved nine galvanometer deflections.

The portion of the specimen to be focused was the same for all wave-lengths and was usually a strip one or two mm. distant from the black-body slit. For each emissivity reading, the zero and sensitivity of the galvanometer were adjusted so as to give maximum accuracy. The observations made with the optical pyrometer were used to calculate the average emissivity for a wave-length of $0\cdot65\,\mu$ by means of the pyrometer correction formula,

$$\frac{1}{T}-\frac{1}{T_{\text{app}}}=-\frac{\log E}{9880}.$$

The galvanometer readings were averaged so as to give for each wave-length a "black-body deflection" and a "specimen deflection". The ratio of these (specimen : black-body) gave the calculated emissivity at this wave-length.

§ 5. CORRECTIONS

One possible source of error, the existence of a temperature difference between the inside and the outside of the hot cylinder, has been investigated by Angell (1911) who found the effect negligible. A second source of error (that of a temperature difference round the cylinder) was studied by Ives (1924) and was also found to be negligible. A photographic examination of the present furnace confirmed this result. It must, however, be remarked that the axial temperature-gradient found with this arrangement was seen, upon micro-photometric examination, to be much larger than that reported by Ives with a specimen of similar dimensions. This error was nullified in the present work by taking the ratio of the radiation from the slit and a parallel section of specimen surface having a similar gradient, but it is probable that the use of such an arrangement as a primary standard of light would be seriously impaired by this gradient unless a much longer cylinder were used.

From the outset of the present work it was found that the emissivity values obtained were rather high not only in the infra-red but also at $0\cdot65\,\mu$. Since the discrepancy appeared to increase with wave-length and was approximately the same for platinum and iron, it pointed to an imperfection in the black-body. The only direct check on this by means of a standard black-body furnace proved to be unobtainable due to war-time difficulties, which made it impossible to obtain and calibrate even a suitable sub-standard. The exact source of the error was not easy to find. The effect of a finite instead of an infinitesimal slit has been discussed by Ives and found to be small provided the slit is sighted from an angle inclined a few degrees in azimuth from the normal. This condition was complied with in the present work and, moreover, the results of Ives were independently confirmed by photographic means. Another investigation by Cunnold and Milford (1934) on the black-body deficiency caused by a finite hole in the wall of a hot tube asserts that in general the error from this source is very small.

A series of experiments with slits ranging from 0·25 to 2·0 mm. in width showed very little variation, so confirming the above. A specially designed specimen having a uniformly heated central portion produced no better result, and hence it was deduced that the axial temperature gradient was not the source of error. Roughening the interior of the cylinder was found to produce no change in the black-body efficiency, doubtless because the increase in emissivity of the interior was being compensated by an increase in diffuse as opposed to specular reflection. Covering the inside of the specimen cylinder with various blacks and oxide coatings was observed also to be ineffective. The provision of hot ends for the cylinder was seen to be impracticable on account of the difficulty of heating these ends separately, and further, because the low heat capacity of the specimen and its high axial temperature-gradient, would have made the formation of a uniform-temperature enclosure almost impossible.

It was also shown that the angle of view and aperture of the observing instrument did not, within reasonable limits, affect the black-body efficiency. It was, however, found that if the cylinder was dented near the electrodes the efficiency of the black-body was increased.

This seemed to indicate that the finite length of the cylinder was responsible for the error since some radiation was being lost at the cold ends of the cylinder. A qualitative study of this may be made by supposing only a finite number of inter-reflections (instead of an infinite number) to take place inside the cylinder. In the case of four inter-reflections, for instance, the radiation from the " black-body " slit is given by

$$EB(1+R+R^2+R^3+R^4)=\frac{EB}{1-R}(1-R^5)=(1-R^5)B,$$

showing a black-body deficiency of R^5. The effect of this on the apparent emissivity may be seen from the following figures for platinum :—

	True E	Calculated deficiency	Apparent E (calc.)
0·65 μ	0·35	11%	0·39
4·0 μ	0·10	60%	0·25

These results are of the order of magnitude found and hence the hypothesis of a finite number of inter-reflections was felt justifiable. To minimize the effect, a special type of specimen was designed in which the middle portion of the cylinder was bulged out into a sphere. It was constructed of platinum by the Baker Platinum Co. The specimen was made in two halves, each consisting of a half cylinder bulged in the middle into a hemisphere. The wall thickness was such that when heated electrically the temperature of the spherical portion was practically uniform. A small slit 1 mm. wide and 1 cm. long on the edge of one hemisphere was used as a black-body source. The emissivity readings obtained with this spherical specimen showed a large improvement over those with the cylinder. The results, indeed, compared so favourably with the available data for platinum, that it was assumed, in default of any conclusive check, that the spherical black-body was sensibly perfect, at least in the near infra-red. It must, however, be pointed out that all absolute readings subsequently derived are liable to be high, the error increasing with wave-length. If these " sphere " results be taken as standard for platinum, then by comparing them with the data obtained for the same metal with the cylindrical specimen, a calibration curve may be drawn to correct " cylinder " values to the " sphere " standard. As an additional argument in favour of the adoption of this standard it may be noted that the results for iron and other metals obtained from the cylinder technique produce values comparing favourably with those of other workers when corrected in this manner. Moreover, extrapolation to the visible region indicates a correction to be applied to Ives' values. This correction is in good agreement with more recent determinations by other methods. Fuller details of the sphere/cylinder calibration are given in the following section.

§ 6. EXPERIMENTAL OBSERVATIONS

Platinum. All the specimens used were of 99·8 % purity platinum obtained from the Baker Platinum Co. In all, six full sets of observations were made with the cylindrical type of specimen and eleven with the spherical specimen. The agreement between different sets is of the order of accuracy of 1% for trend and 5% for absolute value. The average of each collection of data was taken. Smoothed values have been used in figure 2. It must be noted that since most of the possible errors tend to increase the observed emissivity, it is likely that a small constant error is introduced by taking the average of many sets of data, thus making all values slightly too high. The values of $E_{0\cdot65}$ deduced from the pyrometer readings are estimated to be accurate to about 4%.

On a number of occasions the cylindrical specimen was photographed by its own radiation. The variation of emissivity with angle of emission was calculated from the intensity distribution across the cylinder. The results were not very accurate, but as a rough figure it may be stated that at a mean wave-length of 0·55 μ there exists a sharp maximum of about 120 % normal emissivity for an angle of emission of about 85°. (Stephens finds a maximum of 124% at 80° for platinum under similar conditions.)

Using the two sets of values obtained for the emissivity of platinum, a correction curve was plotted showing the calculated efficiency of the cylindrical type of black-body (that is, assuming the spherical type to be perfect). Since the two techniques had been used for platinum under approximately the same conditions and over slightly different ranges and intervals it was thought worth while to synthesize the seventeen sets of results. The final figures (smoothed) for the emissivity of platinum at an average temperature of 1125° are as follows:—

Table 1

λ_μ	0·65	1·0	1·1	1·2	1·3	1·4	1·5	1·75	2·0	2·25
E (corr.)	0·330	0·293	0·287	0·284	0·284	0·276	0·270	0·255	0·240	0·228

λ_μ	2·5	2·75	3·0	3·25	3·5	3·75	4·0	4·25	4·5	4·75
E (corr.)	0·218	0·206	0·196	0·188	0·180	0·172	0·165	0·157	0·150	0·145

Because the correction curve must be independent of the nature of the emitting metal and can be a function only of its true emissivity and reflectivity, it is possible to use this curve to correct readings obtained for other metals with the "cylinder" technique. Before this can be done, however, two points must be considered:

(*a*) The curve must be extrapolated at both ends to cover the range of readings found for the other metals to be investigated. Bearing in mind the qualitative theory already given of this type of deficiency, and also the work of Buckley on similar problems, a reasonable extension of the curve was made. This is shown in figure 2, which gives the observed values for platinum together with full curves showing the black-body efficiency (*B*) plotted against (i) the true emissivity (*E* sphere) and (ii) the apparent emissivity (*E* cylinder). The use of this latter curve enabled all experimental results to be corrected by multiplying by the appropriate value of *B*.

(*b*) It will be seen that apart from the full curve shown, the experimental values indicate a secondary component in the correction curve. This subsidiary deficiency has a "cocked hat" form with a maximum of the order of 4%. In view of the fact that this maximum coincides with the maximum in the radiant energy curves for the temperatures used, it is probably due to the scattering of radiation or to a heating effect involving re-radiation from the spectrometer slits, etc. Assuming some explanation of this nature, a correction may easily be made by plotting the extra deficiency (*C*) against wave-length for platinum and applying this to the results for other metals. A graph of this correction is inset in figure 2. In spite of the somewhat arbitrary nature of this second correction it was thought sufficient since the effect of it is only of the same order as the experimental error. In all subsequent results, values have been corrected by multiplying by the appropriate value of *B* and then subtracting a percentage *C*.

Iron. The metal used in all experiments was obtained in the form of 5 mm. diameter electrode rods produced for spectroscopic work by Messrs. Johnson, Matthey. An analysis indicated a purity of 99·96%. The metal was rolled into thin sheet between clean steel rollers. Throughout the observations no deterioration of the surface was observed even after heating in a hydrogen atmosphere for over 12 hours at 1200° c. In all, four different specimens were used and, with these, nine sets of observations were obtained. The observed values were corrected and smoothed graphically to give final figures, the accuracy of which is estimated to be similar to that for platinum. These figures for pure iron at an average temperature of 1245° c. are as follows:—

Table 2

λ_μ	0·65	1·0	1·1	1·2	1·3	1·4	1·5	1·75	2·0
E (cyl.)	0·494	0·397	0·389	0·378	0·368	0·363	0·357	0·341	0·334
E (corr.)	0·437	0·340	0·330	0·316	0·306	0·298	0·290	0·270	0·260
λ_μ	2·25	2·5	2·75	3·0	3·25	3·5	4·0	4·5	
E (cyl.)	0·326	0·323	0·315	0·312	0·306	0·301	0·288	0·282	
E (corr.)	0·252	0·248	0·244	0·240	0·237	0·235	0·225	0·218	

Molybdenum. High purity electrode rods of this metal proved too brittle for rolling into sheet in the same manner as iron. Thin sheet of ordinary chemical purity, obtained from Messrs. Johnson, Matthey was therefore used. An excellent polished surface free from blemishes was easily obtained and maintained throughout the experiments. The metal was heated in a static atmosphere of hydrogen and 16 sets of observations were obtained. These were averaged and corrected from the calibration curves. The mean temperature of the specimens was 1226° c. The average accuracy of each observed point is estimated at 2 %.

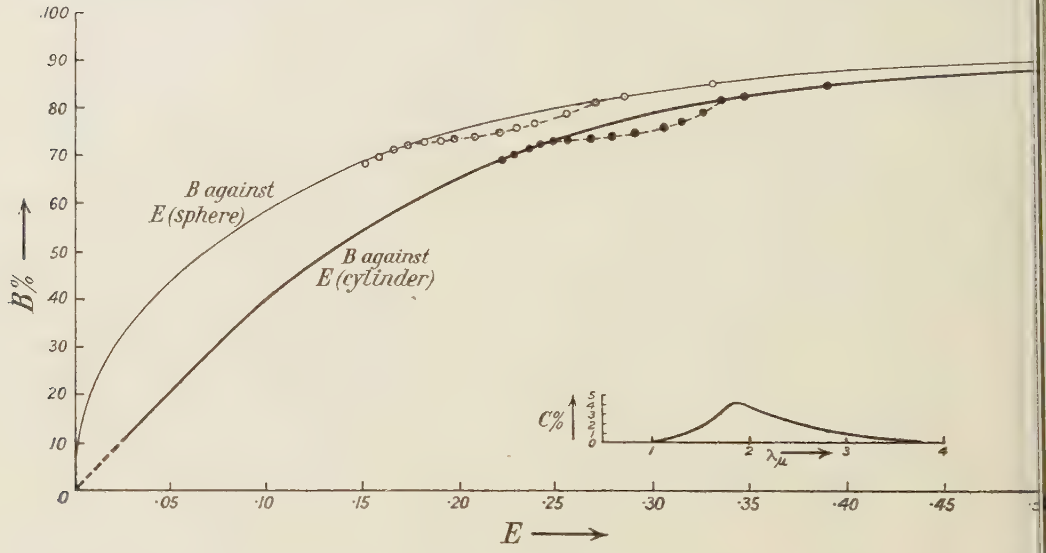


Figure 2. Efficiency of black-body.

Table 3

λ_μ	0·65	1·0	1·1	1·2	1·3	1·4	1·5	1·75	2·0
E (cyl.)	0·360	0·410	0·387	0·372	0·351	0·335	0·319	0·297	0·269
E (corr.)	0·299	0·350	0·327	0·308	0·288	0·270	0·251	0·220	0·197
λ_μ	2·25	2·5	2·75	3·0	3·25	3·5	3·75	4·0	4·5
E (cyl.)	0·264	0·243	0·242	0·223	0·214	0·208	0·200	0·193	0·165
E (corr.)	0·185	0·170	0·163	0·151	0·142	0·135	0·129	0·122	0·095?

Copper. High grade copper of "electrical" purity was used in all observations. Some difficulty was experienced in working with this metal, both on account of its low melting point and also because of its sudden fusion, caused no doubt by recrystallization taking place at high temperatures. Readings had to be restricted to the range 1·5 to 4·0 μ on account of the small quantity of energy emitted at the low working temperature (901° c.). Ten sets of observations were obtained and averaged. The surface appeared bright and uncontaminated by impurity, but in spite of this the average error was somewhat high (5–10 %). Another source of error is the extrapolation needed in the correction curve to permit its use in connection with the low emissivity of copper, and for these reasons the results for this metal may be rather unsatisfactory.

Table 4

λ_μ	0·65	1·5	2·0	2·5	3·0	3·5	4·0
E (cyl.)	0·420	0·150	0·141	0·118	0·105	0·102	0·088
E (corr.)	0·360	0·079	0·065	0·052	0·043	0·038	0·032

Nickel. As with iron, the metal was obtained from Messrs. Johnson, Matthey in the form of rods produced for spectroscopic purposes and estimated to contain 99·97 % nickel. The rod was rolled into thin sheet between clean steel rollers. No trouble was found in producing and maintaining a bright, clean surface, although buckling of the specimen was more pronounced with this metal than with any other used. Because of this a number of the sets had to be abandoned and discarded. In spite of this twelve complete runs of observations were made and the results averaged. These readings were corrected from the curve. The emissivity of pure nickel at an average temperature of 1110° c. was found to be:—

Table 5

λ_μ	0·65	1·0	1·1	1·2	1·3	1·4	1·5
E (cyl.)	0·600	0·358	0·373	0·358	0·345	0·336	0·319
E (corr.)	0·540	0·292	0·314	0·292	0·280	0·270	0·250

λ_μ	2·0	2·5	3·0	3·5	4·0
E (cyl.)	0·296	0·279	0·256	0·241	0·231
E (corr.)	0·220	0·205	0·187	0·174	0·162

§ 7. CRITICAL SUMMARY OF EXPERIMENTAL OBSERVATIONS

For convenience the final, corrected and smoothed values for the high-temperature emissivities of the five metals investigated have been tabulated together. The figures enclosed in brackets are doubtful, but apart from these a mean deviation from the smooth curve of the order of 0·5 % for the emissivity values can be claimed (with the reservations made in the section on black-body efficiency).

Table 6. Summary of data (emissivities, %)

λ_μ	Pt 1125° c.	Fe 1245° c.	Cu 901° c.	Mo 1226° c.	Ni 1110° c.
0·65	22·0	43·7			(54·0)
1·0	29·3	34·0		35·0	(29·2)
1·1	28·7	33·0		32·7	31·4
1·2	28·4	31·6		30·8	29·2
1·3	28·4	30·6		28·8	28·0
1·4	27·6	29·8		27·0	27·0
1·5	27·0	29·0	7·9	25·1	25·0
1·75	25·5	27·0		22·0	
2·0	24·0	26·0	6·5	19·7	22·0
2·25	22·8	25·2		18·5	
2·5	21·8	24·8	5·2	17·0	20·5
2·75	20·6	24·4		16·3	
3·0	19·6	24·0	4·3	15·1	18·7
3·25	18·8	23·7		14·2	
3·5	18·0	23·5	3·8	13·5	17·4
3·75	17·2			12·9	
4·0	16·5	22·5	3·2	12·2	16·2
4·25	15·7				
4·5	15·0	21·8		(9·5 ?)	
4·75	14·5				

From this table we find:—

Platinum. The anomaly giving a minimum at 1·1 μ and a maximum at 1·2 μ occurred in all series taken in this region and, therefore, is unlikely to be spurious. In support of this it may be noted that McCauley (1913) found a similar maximum at 1·1 μ for platinum at 1300° c. The size of the anomaly found in the present work is less than that noted by McCauley but against this is the fact that according to his work the maximum is reduced at 1000° c. to a point of inflexion. Although the present work is in good agreement with McCauley for $\lambda > 3\,\mu$, it becomes increasingly low at short wave-lengths and eventually approximates in the visible region to the room-temperature data of Hagen and Rubens (1903) and Coblentz (1911). Incidentally, the data of these workers and of Försterling and Freedericksz (1913) all show fluctuations from a smooth curve in the interval 0·9 to 1·5 μ, although in no case have sufficient points been taken to show that this is not an example of random error. The value of $E_{0\cdot 65}$ is in excellent agreement with that given by Foote, Fairchild and Harrison (1921). At the other end of the wave-length range a comparison may be had by extrapolating the data given by Hagen and Rubens (1909) for the emissivity at 6·65 μ (using a *Reststrahlen* method). In the range 100° to 500° c. they found $E = 0\cdot 0383\ (1 + 0\cdot 00162T)$. This formula gives $E = 0\cdot 108$ for 1125° c., this point falling on the present curve. Since the curve lies wholly above the corresponding room-temperature values, it will be seen that there can exist no real X-point in the range covered. This conflicts with experiments carried out by Hagen and Rubens (1910) at 2, 4, and 6 μ between 635° and 1455° c. Their results, which are, unfortunately, on an arbitrary scale, show that at 4 μ and 6 μ the Hagen–Rubens relation is valid, but for 2 μ a negative temperature coefficient occurs indicating an X-point at about 2·15 μ. On the basis of the present work and by the comparison which has been made with other data for short wave-lengths, it is believed that these measurements of Hagen and Rubens are erroneous.

Deficiency correction. An interesting application of the deficiency curve found experimentally for a cylinder "black-body" is the correction of Ives's (1924) standard of visible radiation. Ives found the brightness of his standard (of the same type and size as the present specimen) to be 55·40 candles/cm^2 at the freezing point of platinum. For the visible region the mean emissivity of platinum at 1773° c. is approximately 0·35, for which the extrapolated calibration curve indicates a "black-body" efficiency of about 86 %. This increases the figure given by Ives to 64·5 cp./cm^2, which is in better agreement with the more modern determination of 61·0 cp./cm^2 Moreover, Ives reported that the colour temperature of his source was unaccountably 35° c. too high. A graphical computation involving the visibility function and estimated "black-body" deficiencies at different wave-lengths showed that such a figure as 30–40% error in the colour temperature should reasonably occur.

Iron. The observed curve was free from any anomalies. No previous measurements have been made on the pure metal in this region of the spectrum at high or low temperatures. Indeed, the only high-temperature results existing for steel are due to Hagen and Rubens (1910). They cover, roughly speaking, the range 0–300° c. and 0·78–5·0 μ, 6·65 μ, 8·85 μ and 25·5 μ. Since the steel used had a very high nickel content it is futile to attempt any quantitative comparison of results. The available figures for reflectivity (at room temperature) are two sets by Hagen and Rubens ("ungehärtet Stahl"), one by Coblentz (iron of unspecified purity) and one by Ingersoll (1910) (tool steel). All these curves intersect the present high-temperature curve at points ranging from 1·0 to 1·5 μ. This is in qualitative agreement with the high-temperature observations of Hagen and Rubens which indicate an X-point in the neighbourhood of 1·0 μ.

Molybdenum. The only previous data existing for the infra-red emissivity of this metal are those derived from the room-temperature reflectivity figures given by Coblentz (1911), and the only high-temperature values are those due to Whitney (0·67 μ) and to Worthing (0·47 μ) and (0·67 μ). The present curve (which shows no anomalies) has the same trend as that given by Coblentz but a point of intersection occurs in the region of 1·4 μ. Below this point the temperature coefficient is negative and above it positive. The existence of such a negative coefficient is in agreement with Worthing's (1925) investigation which showed that at 0·47 μ the emissivity decreased from about 0·42 at 400° K. to 0·36 at 2800° K. Similarly at 0·67 μ the decrease was from 0·42 at 400° K. to 0·33 at 2800° K. This is, however, at variance with the finding of Whitney (1935) who reported that between the temperatures of 1400° K. and 2000° K. the emissivity at 0·67 μ was constant. It is to be suspected that the small variation was masked by experimental error or else that the properties of a specimen subjected to heat-treatment and "baking-out" are different from those of the metal used, both in the present work and in that of Worthing. The present value for $E_{0\cdot65}$ is in good agreement with Worthing's and fits well on the extended curve for the infra-red region.

Copper. As has been previously noted, the low emissivity of this meta considerably reduces the accuracy of the present method. In spite of thi the values have the same trend and order of magnitude as those found b Hurst (1933), to whom the only previous infra-red measurements at high temperatures are due. The present curve intersects that of Hurst in the regio of 2·9 μ. At 1·5 μ it is 2% higher and at 4 μ it is about 1% lower than Hurst' values. Infra-red measurements at room temperature have been made b Försterling and Freedericksz (1913), by Hagen and Rubens (1903) and by Ingerso (1910). The present curve lies well above all these, indicating the absence o any X-point in this region. This is in agreement with Bidwell (1913) wh finds a small positive temperature coefficient for both the solid and liquid meta at 0·6 μ, but in disagreement with Burgess (1909) who measured a small negativ coefficient for the liquid metal in this wave-length region. Schubert (1937 has also investigated the reflectivity of the hot metals but his finding is that n true coefficient could be detected within the experimental limits.

Nickel. The present curve is uniformly 5 % higher than the figures foun by Hurst under similar conditions. It lies wholly above the room-temperatur curve of Hagen and Rubens (1902–3) and hence no X-point is found. It ha already been noted that buckling of the specimen was rather troublesome in th case of nickel, and this may well have resulted in a larger black-body deficienc and hence to the 5 % discrepancy with Hurst's results. This is made mor likely since a number of workers have noted an X-point for this metal (e.g. Hurs $\lambda_x = 1{\cdot}85\,\mu$, Cennamo (1939) $\lambda_x = 2{\cdot}25\,\mu$, Rubens (1910) $\lambda_x = 1{\cdot}4\,\mu$, Reid $\lambda_x = 2{\cdot}1\,\mu$) This is supported by the work of Bidwell, who finds a negative temperatur coefficient in the visible, but in opposition Wahlin and Wright (1942) find tha with a properly baked-out specimen the emissivity is sensibly constant (c Whitney's result for copper). The value obtained for $E_{0{\cdot}65}$ is a reasonabl continuation of the present curve, but again this result is too high to give a temperature coefficient of the order of that obtained by other workers. It mus therefore be presumed that this curve is about 5 % too high and that consequentl an X-point may occur in the region of 2·0 μ.

ACKNOWLEDGMENTS

This research forms part of a programme being undertaken for the Foundr Steel Temperature Sub-Committee of the Steel Castings Research Committe of the British Iron and Steel Research Association and Iron and Steel Institute with their financial support.

In conclusion I wish to thank Dr. H. Lowery, Principal of the S.W. Esse Technical College, for his direction of the research.

REFERENCES

Angell, 1911. *Phys. Rev.*, **33**, 422.
Beekman and Oudt, 1925. *Z. Phys.*, **33**, 831.
Bidwell, 1913. *Phys. Rev.*, **1**, 482; 1915. *Ibid.*, **3**, 439.
Buckley, 1927. *Phil. Mag.*, **4**, 753; 1928. *Ibid.*, **6**, 447; 1934. *Ibid.*, **17**, 576.
Burgess, 1909. *Bull. Bur. Stand.*, **6**, 111.
Cennamo, 1939. *Nuovo Cimento*, **16**, 253.
Coblentz, 1908. *Bull. Bur. Stand.*, **5**, 339; 1911. *Ibid.*, **7**, 197; 1918. *Ibid.*, **5**, 31

CUNNOLD and MILFORD, 1934. *Phil. Mag.*, **18**, 561.
FOOTE, FAIRCHILD and HARRISON, 1921. *Sci. Pap. Bur. Stand.*, 170.
FÖRSTERLING and FREEDERICKSZ, 1913. *Ann. Phys., Lpz.*, **40**, 201.
HAGEN and RUBENS, 1900. *Ann. Phys., Lpz.*, **1**, 352; 1902 a. *Ibid.*, **8**, 1; 1902 b. *Ibid.*, **8**, 432; 1903. *Ibid.*, **11**, 873; 1909. *Abh. Preuss. Akad. Wiss.*, **16**, 478; 1910. *Ibid.*, **23**, 467.
HURST, 1933. *Proc. Roy. Soc.*, A, **142**, 466.
INGERSOLL, 1910. *Astrophys. J.*, **32**, 265.
IVES, 1924. *J. Franklin Inst.*, **197**, 147.
LANGMUIR, 1916. *Phys. Rev.*, **7**, 302.
MCCAULEY, 1913. *Astrophys. J.*, **37**, 164.
MENDENHALL, 1911. *Astrophys. J.*, **33**, 91.
PRICE and LOWERY, 1944. *J. Iron and Steel Inst.* (Steel Castings Research Committee), **149**, 523.
REID, 1941. *Phys. Rev.*, **59**, 161.
RUBENS, 1910. *Phys. Z.*, **11**, 139.
SCHUBERT, 1937. *Ann. Phys., Lpz.*, **29**, 473.
STEPHENS, 1936. *J. Opt. Soc.*, **29**, 158.
WAHLIN and WRIGHT, 1942. *J. Appl. Phys.*, **13**, 40.
WHITNEY, 1935. *Phys. Rev.*, **48**, 458.
WORTHING, 1925. *Phys. Rev.*, **25**, 846; 1926. *Ibid.*, **28**, 174.
WORTHING, 1940. *J. Appl. Phys.*, **11**, 421.

THE TEMPERATURE VARIATION OF THE EMISSIVITY OF METALS IN THE NEAR INFRA-RED

BY DEREK J. PRICE,
S.W. Essex Technical College, London

Communicated by H. Lowery; MS. received 15 *July* 1946

ABSTRACT. Evidence for the existence of a wave-length, peculiar to each metal, for which the temperature coefficient of emissivity is zero is analysed and commented upon. Further experimental data obtained from the author's measurements of infra-red emissivity of hot metals are submitted in support, and it is shown that the appearance of such an *X-point* phenomenon would explain many of the divergent results obtained for the temperature variation of emissivity in the visible region, besides providing a new series of characteristic wave-lengths for metals.

THE Hagen and Rubens (1903) approximation has been tested experimentally for a number of metals and has been shown to account accurately for the emissivity and the temperature coefficient of emissivity in all cases for wave-lengths greater than about 10μ. For shorter wave-lengths, experimental results diverge from this theory until, in the visible region, the Hagen–Rubens approximation is found to yield no qualitative picture of either emissivity or its temperature variation. The position is further obscured by the fact that while the experimentally observed temperature coefficient of emissivity in the visible is always much smaller than that demanded by this formula, the data show large discrepancies between different workers. As an example may be cited a collection of data given by Worthing (1941) for the case of platinum (figure 1). The effect sought i certainly much smaller than that expected from the Hagen–Rubens formula indeed it is to be presumed that just because it is so minute, the existence of an small methodical error or a slight change of surface could easily produce just th discrepancy noted in these results. The appearance of conflicting data is no confined to the case of platinum, but is to be found in nearly all investigations o temperature variation of the emissivity, reflectivity, and optical constants of metal in the visible region.

A most interesting effect, contrary to all established notions, has been found b many workers who have observed a *decrease* in emissivity with rise in temperature The existence of such a negative temperature coefficient in the visible region mus imply that somewhere between it and the region of validity of the Hagen–Ruben relationship (where the temperature coefficient must be positive) there is a wave length where the emissivity is constant with temperature, that is, the coefficient i zero. This point, here for shortness termed the *X-point*, has been observe

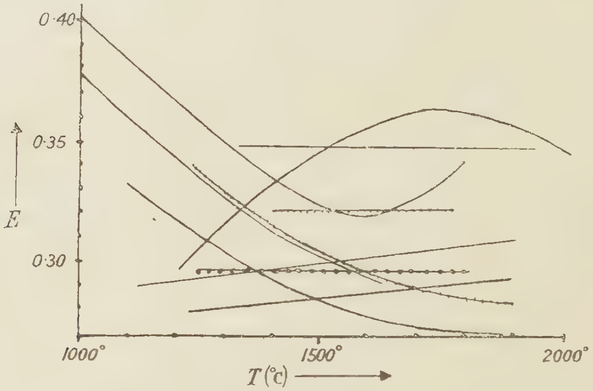


Figure 1. $E_{0\cdot6\,\mu}$ for platinum (adapted from Worthing, 1941, p. 1180).

directly in at least four investigations made on the temperature variation of reflectivity of metal surfaces. (Since, by Kirchhoff's Law, the sum of the reflectivit and emissivity of a non-transparent medium must be unity, these results may b used to calculate emissivity values). The data given by these four investigation are by no means in complete accordance. For instance, Rubens (1910) gives a X-point for nickel at 1·25 μ whereas Reid (1941) finds it at 2·15 μ, and in the case o tungsten, Weniger and Pfund (1919) note the X-point at 1·27 μ and Ornstei (1936) (collating the observations of Hamaker, Vermeulen and van Zelst) finds it a 1·64 μ. The effect is so small, and the disturbance produced by heating a polishe metal surface so large, that the exact location of an X-point by these direct mean is seen to be a matter of great difficulty, and it is therefore hardly surprising to fin the discrepancies noted.

Both Rubens (for nickel) and Ornstein (for tungsten) find the emissivity t increase linearly with temperature. The data of the latter worker are ver striking, for when they are plotted, it can be seen that all the curves of emissivit against wave-length at temperatures ranging from 300° K. to 3000° K. intersec accurately at one point. Unfortunately it is not clear from the original paper whether measurements in the region of the X-point were actually made at the large number of temperatures quoted, or whether these figures have been interpolated from readings taken at the two limiting temperatures only. Certainly an X-point does exist here, and, if the results have not been obtained by interpolation, it would appear that this wave-length is a constant property of the metal over an unusually great range of temperature. In any case, the existence of a specific wave-length in the near infra-red region, characteristic of each metal, is sufficiently interesting to merit further examination, even if this X-point is not absolutely constant at all temperatures.

Since the effect sought appears to be so small as to be almost masked by the experimental errors associated with its direct observation, it is necessary to resort to an indirect attack . Data found for high temperatures by an emissivity technique may be compared with data found for room temperatures from the measurement of reflectivity, and from this information the temperature coefficient may be calculated and the X-point (if any) detected and located. Two main criticisms may be made of this method :—

1. It is necessary to assume a linear variation of emissivity with temperature in order to calculate the coefficient. Since, however, the effect is small and also any deviation from a linear relation would not influence the location of an X-point, this criticism may be neglected.
2. Different experimental techniques and even different specimens have been used for the two temperatures. Against this it may be said that only figures obtained for the pure metal surface should be used, and whenever possible the data compared should be the mean of a number of independent investigations.

In the case of nearly all the common metals there exist sufficient data for room temperatures, e.g. those obtained in the classical investigations of Coblentz and of Hagen and Rubens, and hence wherever experiments have been made at high temperatures it is possible to carry out the above comparison.

Leaving for the moment a number of points of intersection occurring in the visible region, the following X-points obtained by this method may be added to the four previously noted :—

Author for high-temperature results	Metal	X-point (μ)
Hurst, 1933	Cu	2·2
Hagen and Rubens, 1910	Fe	1·0
Hurst, 1933	Ni	1·8
Cennamo, 1939	Ni	2·5
McCauley, 1913	Pd	1·0
Davisson and Weeks, 1924	W	1·28
Forsythe and Worthing, 1925	W	1·3

Again, it is found that the agreement is not very good, due doubtless to the partial masking of the effect sought by the experimental and systematic errors. In the transition region between this X-point and the limit of validity of the Hagen–Rubens relationship, the temperature coefficient is larger and is no longer masked by the experimental inaccuracies. Data obtained by the present author (1947) for the infra-red emissivity at high temperatures of a number of metals enables the comparison method described above to be used for determining the variation o the temperature coefficient of emissivity with wave-length. The room-temperature values employed in this comparison have all been obtained from a smoot curve drawn through the mean of all available results for each metal in the ultra-violet, visible, and infra-red regions.

Although a complete examination of the temperature coefficient, obtained fo each wave-length by measuring the emissivity at a number of temperatures would be more desirable than the room-temperature and single high-temperatur here employed, this plan was found impossible. The available range (*circa* 900 °c. to the melting point) was too small, and the systematic errors inheren in the black-body calibration and comparison were too large to enable useful o consistent figures to be obtained for the coefficient, using an emissivity techniqu of this nature.

The existence of a complete curve for the variation of temperature coefficien with wave-length between the Hagen–Rubens limit and the X-point region, make possible for the first time a fairly accurate location of this X-point to be made b extrapolation of the smooth curve obtained to cut the wave-length axis at the poin of zero temperature coefficient. As will be seen, in some cases, the experimenta curve departs in the X-point region from the trend found for longer wave-lengths This may be either a real phenomenon due to the proximity of the absorptio bands of the bound electrons, or it may in some measure be due to the experimenta and systematic errors previously noted.

Let us examine in more detail the nature and trend of these temperature coefficient curves (figure 2*a*). Starting at the long wave-length end, we see tha above the limit of validity of the Hagen–Rubens relationship the temperature coefficient of emissivity must be half the temperature coefficient of electrica resistivity of the metal. This follows from the fact that in this region emissivity is proportional to the square root of resistivity. As has been stated, the validity of this law has been conclusively established in the now classical work of Hagen and Rubens. Below the limit of validity, the temperature coefficient decreases almost linearly with wave-length up to the region of the X-point. At either end of the linear region the curve turns, at the upper end to meet the horizontal limi predicted by the Hagen–Rubens theory, and at the lower to come eventually to the origin. This latter point follows from the fact that the emissivity of all bodies must be unity at the limit of zero wave-length, and hence the temperature coefficient must be zero.

The temperature coefficient graphs deduced, as stated, from the present author's measurements on molybdenum, iron and nickel are shown in figures 2*c*, 2*d* and 2*e*. That for platinum has been given elsewhere (Price, 1946). The data for copper were somewhat inaccurate, due to the low temperature at which results had to be taken, and the only conclusion which could be drawn was that within the experimental error the temperature coefficient was of the order predicted by the Hagen–Rubens relationship for wave-lengths beyond 1·5 μ, i.e. the horizontal portion of the curve extended to much shorter wave-lengths than for any of the other metals investigated. The figures for molybdenum are very striking since they show clearly all the phenomena noted, an X-point at 1·8 μ, the turning round to the Hagen–Rubens limit starting point at about 2·5–3·0 μ, and the appearance of a minimum at 1·0 μ where the curve turns to go through the origin Iron has an X-point at 1·55 μ, but the turning of the curve at either end is only just indicated in the range covered by observation. The curve for nickel is again too short to show the turning, but besides this, the rather surprising fact is noticed that the temperature coefficient is positive throughout. This omission of the X-point noted by Rubens, Reid, Cennamo, and Hurst, is possibly due to a discontinuity at the Curie point vitiating the assumption of a linear increase with temperature from 0° c. to 1100° c. In this case an investigation over a range of temperatures using a more suitable technique (e.g. reflectivity) would be very desirable.

The existence of an X-point, whether real or virtual, is seen to provide a key both to the short-wave region and to the transition period. In the former case, between the X-point and the origin, the temperature coefficient is small and the curve may show considerable fine structure in the region of absorption bands. Data for tungsten calculated from Ornstein's values (figure 2*b*) indicate the sort of fine structure to be found. In some cases it is possible that one or more of these small maxima or minima should intersect the wave-length axis, so giving one or more further points of zero temperature coefficient. These "spurious X-points" occurring in the region of a band of selective reflection have been observed by Fujioka and Wada (1934) (Ag, 0·31 μ, Au, 0·47 μ, Cu, 0·5 μ), by Worthing (1921) (Au, 0·47 μ), by Ebeling (1925) (Cu, 0·5 μ), and by Ornstein and Van der Veen (1939) (Fe, 0·45 μ ?). It is, of course, somewhat difficult to distinguish between the real and the spurious X-points. Indeed, the position may be said to be satisfactory only when a complete curve of temperature coefficient against wave-length is known. In the cases quoted, however, the presence of a known absorption band at the stated wave-length makes it evident that the phenomenon is due to the temperature variation of this band. This type of explanation cannot serve for the X-point values previously quoted, since there seem to be no known strong resonance frequencies for metals in the region 1·0–2·5 μ.

The only comparative account of this X-point phenomenon is due to Worthing (1926) who, in a short paragraph, quotes his own result for gold ($\lambda_x = 0{\cdot}46$) and that of Weniger and Pfund (1919) for tungsten ($\lambda_x = 1{\cdot}27$), and from these values conjectures that the X-point wave-lengths may be proportional to the melting points of the elements (Au = 1336° K., W = 3655° K.). This is at variance with the fact that λ_x for nickel is of the same order as that for tungsten and in addition it is felt that the divergence of the values given by different workers is too high to allow any law to be formulated on the basis of just two random results. Further it is possible that the figure quoted for gold represents a "spurious" X-point due to the influence of the absorption band in the visible region.

There are at least two possible explanations of the X-point phenomenon, but in neither case is the associated theory sufficiently developed to provide a more exact picture. It is well known that, in general, an absorption band shifts towards higher wave-lengths and becomes broader with increase in temperature. It has also been noted by Schaum and Wustenfield (1911) that the broadening is more pronounced on the long wave side of the band. This type of change is shown in figure 3, and is seen to entail an intersection of the two curves, i.e. an X-point.

It is possible that a similar type of explanation should be applicable to the free-electron portion of the metal. This would mean that the half-width of the

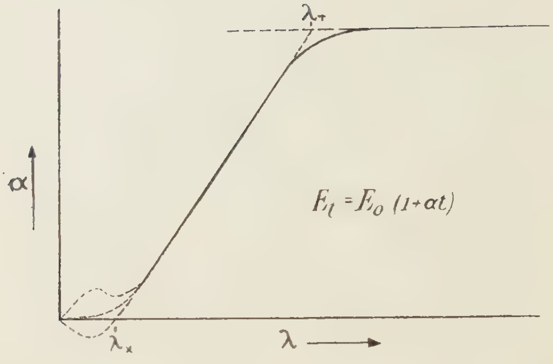


Figure 2 *a*. Idealized curve of temperature coefficient of emissivity against wave-length (cf. Price, 1946).

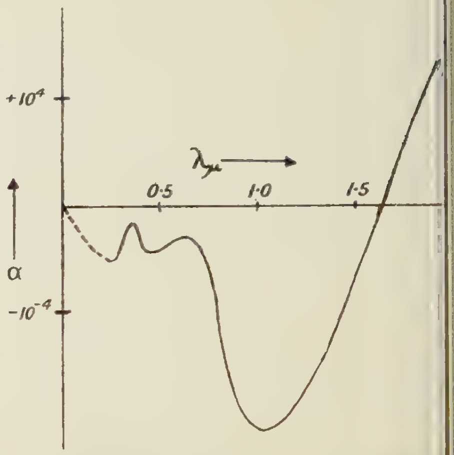


Figure 2 *b*. Temperature coefficient tungsten at 1600° K. (after Ornstein, 1

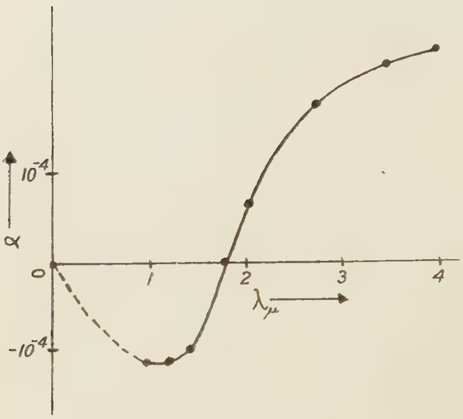


Figure 2 *c*. Temperature coefficient for molybdenum.

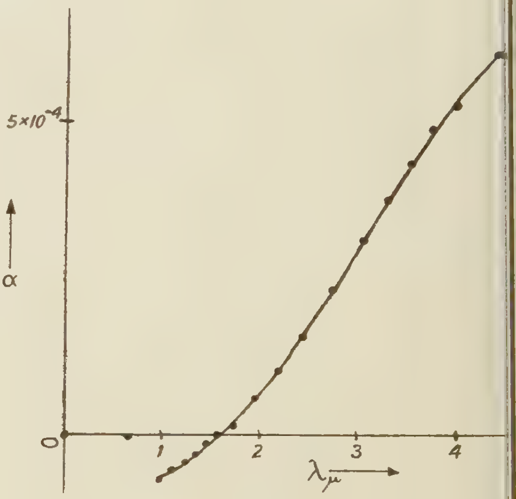


Figure 2 *d*. Temperature coefficient for iro

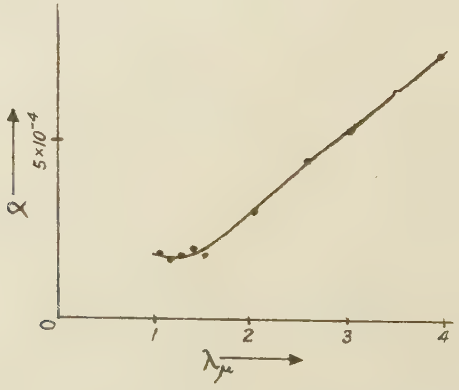


Figure 2 *e*. Temperature coefficient for nickel.

absorption band of the free electrons should decrease with increase of temperature while the long wave-length values of emissivity should, in accordance with the Hagen–Rubens relationship, increase with temperature. This type of explanation is illustrated diagrammatically in figure 4 (*a*). The objection to this explanation is that usually the width of an absorption band increases rather than decreases with rise in temperature.

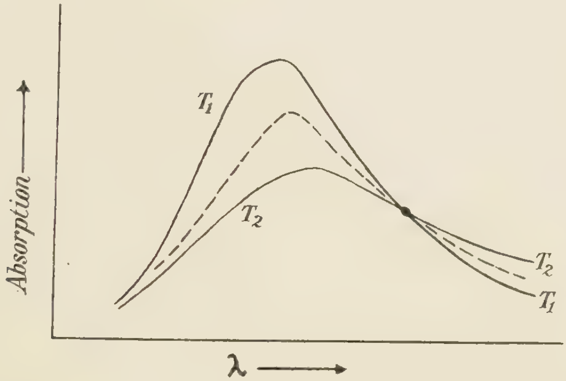


Figure 3. Influence of rise of temperature on an absorption band.

An alternative explanation is that the X-point represents a balance between two opposing effects due on the one hand to the free-electron contribution, and on the other to a bound-electron absorption band. This type of effect is shown diagrammatically in figure 4 (*b*), where it may be seen that the rise in emissivity due to the free-electron portion is balanced at the X-point by the broadening of the band due to one class of bound electrons. It is probably this effect which produces the "spurious" X-points previously noted in the region of the absorption bands at 0·25–0·6 μ. As has been said, however, the absence of strong resonance wave-lengths in the region 1·0–2·5 μ makes this explanation untenable in the case of the true X-point. Now although no selective effects have been found at room temperature, a few investigators have noted that at high temperatures a new absorption band appears in the near infra-red in the very region demanded by

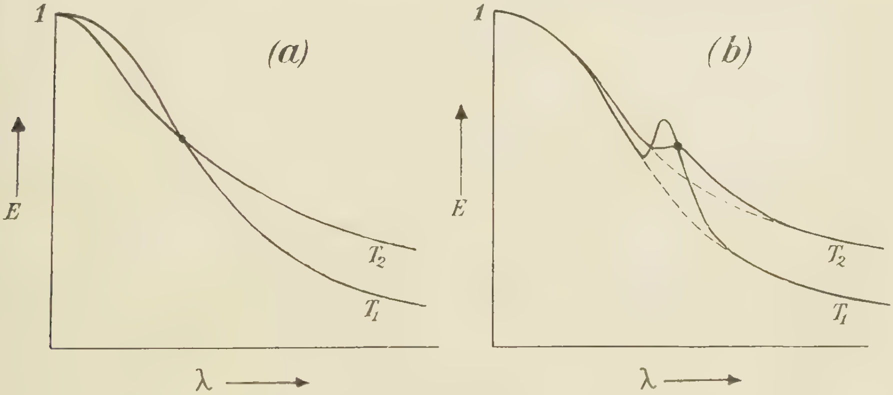


Figure 4. Influence of rise of temperature on emissivity curves.

the X-point. For instance, Ornstein's data (1936) for tungsten show a new maximum developing at about $2{\cdot}0\mu$, and McCauley's investigations (1919) of tantalum, platinum and palladium all show new maxima appearing at high temperatures in the region of $1{\cdot}0\mu$. This result for platinum has been confirmed by the present author, who detected a maximum at $1{\cdot}25\mu$ at 1125° c. In all these cases the maximum and minimum appear to become more pronounced with increase in temperature, and it is quite possible that the appearance of the selective absorption region, hidden at room temperatures, is responsible for the X-point phenomenon. Little more can be said of this possibility without a full catoptric investigation of the optical constants of metals at high temperatures.

Certainly it can be seen that emissivity methods are not sensitive enough to measure accurately the small temperature coefficients existing near the X-point and in the visible region, although some information has been obtained from a series of measurements at longer wave-lengths. This information, in the form of temperature-coefficient graphs, shows clearly the manner in which the Hagen–Rubens relationship breaks down at short wave-lengths. Further, these graphs are characterized by the appearance of an X-point which not only entails the existence of negative temperature coefficients, but also implies that an explanation of the divergent results obtained by many workers may be found if the sensitivity of this X-point to surface contamination, crystal structure, etc. is assumed.

ACKNOWLEDGMENT

I wish to thank Dr. H. Lowery, Principal of the S.W. Essex Technical College, for his direction of the research.

REFERENCES

Cennamo, 1939. *N. Cimento*, **16**, 253.
Davisson and Weeks, 1924. *J. Opt. Soc. Amer.*, **8**, 581.
Ebeling, 1925. *Z. Phys.*, **32**/7, 489.
Forsythe and Worthing, 1925. *Astrophys. J.*, **61**, 146.
Fujioka and Wada, 1934. *Sci. Pap. Inst. Phys. Chem. Res., Tokio*, **25**, 9.
Hagen and Rubens, 1903. *Ann. Phys., Lpz.*, **11**, 873.
Hagen and Rubens, 1910. *Abh. Preuss. Akad Wiss.*, **23**, 467.
Hurst, 1933. *Proc. Roy. Soc.*, A, **142**, 466.
McCauley, 1913. *Astrophys. J.*, **37**, 164.
Ornstein, 1936. *Physica*, **3**, 561.
Ornstein and van der Veen, 1939. *Physica*, **6**, 439.
Price, 1946. *Nature. Lond.*, **157**, 765.
Price, 1947. *Proc. Phys. Soc.*, **59**, 118.
Price and Lowery, 1944. *J. Iron Steel Inst.*, no. 1, 523.
Reid, 1941. *Phys. Rev.*, **59**, 161.
Rubens, 1910. *Phys. Z.*, **11**, 139.
Schaum and Wustenfield, 1911. *Z. Wiss. Photogr.*, **10**, 213.
Weniger and Pfund, 1919. *Phys. Rev.*, **14**, 427.
Worthing, 1921. *J. Franklin Inst.*, **192**, 112.
Worthing, 1926. *Phys. Rev.*, **28**, 174.
Worthing, 1941. *Temperature, Its Measurement and Control in Science and Industry* (American Institute of Physics; Reinhold Publishing Corp.), p. 1164.

THE FREEZING-IN OF NUCLEAR EQUILIBRIUM

By A. R. UBBELOHDE,
Queen's University, Belfast

MS. received 25 *October* 1946

ABSTRACT. Experimental values of abundance ratios of stable isotopes, and of their mass differences, are used in a graphical method of testing how far the reaction velocities for different nuclear transformations occur at comparable rates under the conditions of freezing-in. With the limited data so far available for exact mass differences (up to atomic mass 57, Fe), a single freezing-in parameter of approximately 10^{10} degrees centigrade is found to account for a large proportion of the stable isotopes. However, a different origin, or very different reaction velocities for the nuclear transformations, is indicated for some of the stable isotopes, such as those of lithium.

§1. INTRODUCTION

MANY examples are known of the freezing-in of molecular or chemical equilibrium, and some of the features of such frozen equilibria may be transposed to the problem of the freezing-in of nuclear equilibria.

Freezing-in occurs as the temperature or density of a system in equilibrium is progressively lowered, when the velocity with which changes of concentration can occur eventually becomes insufficient to adjust these concentrations to the fall in temperature, or density. If the only change occurring were the fall in temperature, Le Chatelier's principle indicates that the composition would lag behind that for true equilibrium, in favour of the endothermic components of the system. But in the case of nuclear equilibrium, freezing-in may have been determined by both temperature and partial pressure changes, and the characterization of freezing-in conditions is more complex. In the discussion which follows, the treatment is simplified deliberately, since no more detail seems to be justified with the experimental information available at present.

Assuming that the various atomic nuclei in the earth's substance were at some stage in equilibrium, this would call for detailed balancing between the various processes of nuclear transformation, so long as equilibrium was maintained. Processes involving a change in atomic number would include those verified in the laboratory, such as

$$ {}^{m}_{n}A + {}^{1}_{1}p \rightarrow {}^{m+1}_{n+1}A \qquad \ldots\ldots(1)$$

where ${}^{m}_{n}A$ is an atom of mass number m and charge n, and ${}^{1}_{1}p$ is a proton. Processes involving a change of atomic mass without change of atomic number would have included transformations such as

$$ {}^{m}_{n}A + {}^{1}_{0}n \rightarrow {}^{m+1}_{n}A \qquad \ldots\ldots(2)$$

where ${}^{1}_{0}n$ is the neutron. But other processes not yet discovered in the laboratory may also have been involved, and one of the purposes of what follows is to examine in what cases such processes may have been operative, and thus to throw fresh light on nuclear transformations.

When the statistical thermodynamics controlling equilibrium in a system i known, the concentrations reached in frozen-in equilibrium may be used t calculate a nominal "freezing-in temperature". This will have a value lyin somewhere in the region where the rate of adjustment of concentrations i beginning to lag behind the rate of change of external conditions. In fact, such freezing-in temperature may be regarded as a parameter characterizing th reaction velocities in the system under the conditions prevailing at the time.

Experimental data for the calculation of such parameters include (i) th relative abundance of elements of different atomic number, first discussed b Harkins (1917); and (ii) the relative abundance of isotopes of the same element It does not necessarily follow that the freezing-in temperature for reactions i which the atomic number changes approximates to that for reactions in which onl the atomic mass changes. Strictly speaking, each of the distinct processe involved in the detailed balancing which maintains equilibrium may freeze-in at a different temperature. It is therefore necessary to select one process at a time in comparing the freezing-in conditions for the different nuclei. Owing to the fac that chemical segregation of some of the elements subsequent to the freezing-in may have falsified some of the abundance ratios for nuclei of different atomi number, data of the type (i) will not be used in the present paper, though a simila procedure can be followed, as in the calculations given below, which are restricte to data of the type (ii).

Fairly accurate information has been collected about the abundance ratios o isotopes of the same element (Seaborg, 1944). These may be used to examine the hypothesis that equilibrium between isotopes of the same element was maintaine at one period of the earth's history by a partial pressure of neutrons in the mass, o sufficient concentration and sufficient translational energy to give reactio velocities adequate to maintain equilibrium. Without going into the detail o these reactions at this stage, if the reaction velocities were comparable for the nuclei of different masses, the same freezing-in parameter would be found, *with respect to this reaction* (equation (2) above), for the whole of the periodic system. On the other hand, exceptionally sluggish nuclear reactions of this type would have greater freezing-in parameters, and exceptionally speedy nuclear reactions would have smaller freezing-in parameters (temperature and partial pressures). A complication arising from radioactive nuclei in equilibrium with stable nuclei is discussed below.

§ 2. STATISTICAL CALCULATIONS OF FREEZING-IN PARAMETERS

Although the equilibria between nuclei, photons, and elementary particles at high temperatures may be governed by statistics which differ substantially from those which hold at ordinary temperatures (Wataghin, 1944; Lattes and Wataghin, 1946), many of the uncertainties cancel out in the particular problem examined in this paper. In reversible processes of isotope formation, such as

$${}^{m}_{n}A + {}^{1}_{0}n \rightleftharpoons {}^{m+1}_{n}A + \Delta H_1, \qquad \ldots\ldots(3)$$

$${}^{m+1}_{n}A + {}^{1}_{0}n \rightleftharpoons {}^{m+2}_{n}A + \Delta H_2, \qquad \ldots\ldots(4)$$

equilibrium constants may be defined in the usual way, e.g.

$$K = [{}^{m+1}_{n}A]/[{}^{m}_{n}A] \cdot [n], \qquad \ldots\ldots(5)$$

where the terms in square brackets refer to the thermodynamic activities of the appropriate species. These equilibrium constants will be related to the heat evolved, ΔH_1, etc., according to standard formulae (cf. Ubbelohde, 1937)

$$\log_e K = -\frac{\Delta H_0}{\boldsymbol{k}T} - \frac{1}{\boldsymbol{k}T}\int \Sigma C_p \, dT + \frac{1}{\boldsymbol{k}}\int C_p \frac{dT}{T} + I'. \qquad \ldots\ldots(6)$$

This expression may be simplified, and freed from a number of experimental unknowns, as follows:

In place of $\log_e K$, we may write $-\log_e [n] + \log_e r$, using (5), and taking the ratio of the thermodynamic activities

$$[{}^{m+1}_{n}A]/[{}^{m}_{n}A] = r$$

under the conditions of freezing-in, as given by the ratio of concentrations r in which isotopic nuclei are found to be stabilized under current terrestial conditions. Owing to the close similarity in each pair of nuclei considered, this approximation appears to be justified, even if the thermodynamic activity itself differs largely from the partial pressure, under the conditions of freezing-in.

In place of ΔH_0, the energy which would be absorbed in the nuclear transformation if it occurred at 0° K., the Einstein rest-mass equation gives

$$\Delta H_0 = \boldsymbol{c}^2 [{}^{m+1}_{n}A - {}^{m}_{n}A - m_n], \qquad \ldots\ldots(7)$$

where the terms inside the brackets refer to the atomic masses of the particles. It is not necessary to know the exact mass m_n of the neutron for what follows, but the difference in mass between two isotopes differing by unity in mass number must be known with high accuracy. It may be written ΔA, so that

$$\Delta H_0 = \boldsymbol{c}^2 [\Delta A - m_n]. \qquad \ldots\ldots(8)$$

Using as mass scale for ΔA, ${}^{16}O = 16$, the unit of mass will be $1{\cdot}66 \times 10^{-24}$ (Aston, 1942), and inserting for

$$\boldsymbol{k} = 1{\cdot}379 \times 10^{-16} \text{ c.g.s.},$$

$$\boldsymbol{c}^2 = 9 \times 10^{20} \text{ c.g.s.},$$

$$\Delta H_0/\boldsymbol{k} = 1{\cdot}08 \times 10^{13} [\Delta A - m_n].$$

A quantity which is also required has the value

$$\Delta H_0/2{\cdot}303\boldsymbol{k} = 4{\cdot}7 \times 10^{12} [\Delta A - m_n].$$

At the freezing-in temperatures it seems unlikely that the specific-heat terms will involve anything but translational energy, since nuclear excitation is probably incompatible with freezing-in, and the moment of inertia of the nucleus I can be neglected.

[Writing $\Theta_{\text{rot}} = \boldsymbol{h}^2/8\pi^2 I\boldsymbol{k}$, if the radius of the nucleus is of the order of 5×10^{-13} cm., the characteristic temperature Θ_{rot} for the excitation of rotation is of the order $2{\cdot}5 \times 10^{13}/A°$ C. where A is the mass number.]

Thus ΣC_p for two particles condensing to one will be equal to $\frac{3}{2}\boldsymbol{k}$, i.e.

$$-\frac{1}{\boldsymbol{k}T}\int \Sigma C_p \, dT + \frac{1}{\boldsymbol{k}}\int \Sigma C_p \, dT/T = \tfrac{3}{2} - \tfrac{3}{2}\log_e T.$$

This expression will not be strictly applicable if equilibrium was frozen-in at high densities and very high temperatures, but it is quite adequate for testing whether the same freezing-in parameters apply to the different atoms of the periodic system.

By a standard result, the constant $I' = \Sigma i_m$, where the constants i_m correspond to the entropy constants for the various particles

$$i_m = \log_e \frac{(2\pi m \boldsymbol{k})^{3/2} \boldsymbol{k} g_m}{\boldsymbol{h}^3}.$$

Apart from constants, i_m is a function of the mass and the spin multiplicity g_m of the nucleus. For the simple equilibrium considered,

$$I' = \log g_{m+1}/g_m + \log C$$

where the term $\log C$ depends only on the neutron, and universal constants, apart from acorrection for the fact that the ratio of the atomic masses $\log ({}^{m+1}A/{}^{m}A)$ differs from unity. Except for hydrogen, this correction can be neglected compared with the other uncertainties still present in the expression.

Inserting these simplifications into (6), rearranging, and using common logarithms, the freezing-in parameters may be calculated from the equation

$$\log_{10} r = -4{\cdot}7 \times 10^{12} \Delta A/T + \log_{10} g_{m+1}/g_m + K_e \qquad \ldots\ldots(7)$$

where the term

$$K_e = 2{\cdot}303 \log_{10}[n] + 2{\cdot}303[-1{\cdot}08 \times 10^{13} m_n/T + \tfrac{3}{2} - \tfrac{3}{2} \log_e T] + \log_{10} C$$

is independent of the masses ${}^{m+1}_{\;\;n}A$, ${}^{m}_{n}A$, and is constant for all nuclei with the same freezing-in parameters T and $\log_{10}[n]$.

The constancy of these parameters may conveniently be tested graphically, by plotting experimental values of $-\log_{10} r$ against experimental values of ΔA. At present, these are available with sufficient accuracy in the case of a limited number of atoms only (Pollard, 1940; Okuda *et al.*, 1941).

In this plot, reproduced in figure 1, it is convenient to include data for isotopes differing by mass number 2 (triangles), as the intermediate isotope is sometimes rare. Data for isotopes differing by mass number 1 are expressed as circles. The thermodynamic equation for the mass difference 2 is readily obtained by adding the equations for the two steps (3) and (4) above, and making the appropriate adjustments to the constants. Provisionally, until more general information is available about nuclear spins, no allowance has been made in the plot for the ratio g_{m+1}/g_m being different from unity in some cases. If j_m is the spin of the isotope of mass number m, the order of correction to $\log_{10} r$ is $\log_{10} \dfrac{(2j_{m+1}+1)}{(2j_m+1)}$, which will normally be considerably less than $\pm 1{\cdot}00$, and so does not obscure the main conclusions from the diagram. Examination of this shows that there is an undoubted general correlation between the mass (or energy) change on forming one isotope from another and the abundance ratio as at present stabilized in the earth's substance.

On the basis of equation (7) above, the freezing-in temperature which corresponds with the heavy line drawn in the diagram has a value given by

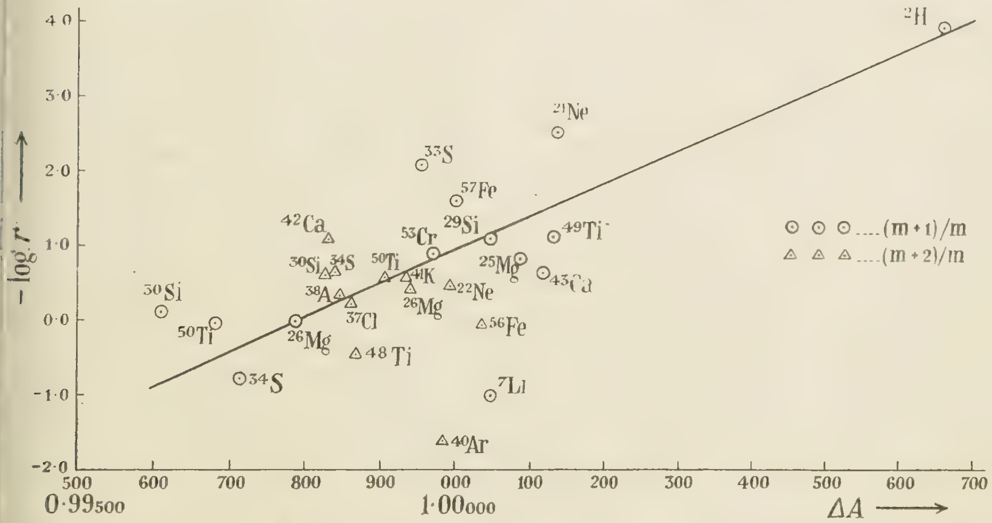
$$\Delta(\log r)/\Delta A \simeq 1{\cdot}0/0{\cdot}0023 \doteq 4{\cdot}7 \times 10^{12}/T$$

or

$$T \text{ (freezing-in)} = 10^{10}\ {}^{\circ}\text{C}.$$

The fact that this freezing-in temperature appears to hold within fairly close limits for a considerable number of atomic species is perhaps rather surprising. It suggests that the reason for freezing-in may have been a falling off in the term $\log [n]$ for the partial pressure of neutrons, which would affect many of the reactions in the same way, rather than inadequate activation energy of the neutrons, which would be expected to occur rather sooner for the larger values of ΔA.

Particular interest is attached to the exceptional values of the freezing-in parameters. At the present stage of our information about nuclear spins it is probably advisable to treat values of $\log r$ within $\pm 1{\cdot}00$ of the freezing-in curve as not yet clearly differentiated. If lines parallel to the curve are drawn above and below it, at this distance, this leaves certain isotopes of Si, Ca, S, and Ne with exceptionally high values of $\log r$, and of Ti, Fe, Ca and Li, Ar with exceptionally low values of $\log r$. The interpretation of these divergences is of considerable importance. This paper will deal with one or two points only:



(i) A trivial explanation for some of the divergences is that experimental values of r, or more probably of ΔA, are incorrect. Further work on accurate measurements of mass differences should clear this up in due course.

(ii) Under the conditions of freezing-in, an appreciable proportion of the nuclear species in the earth's mass may have been radioactive, i.e. capable of undergoing spontaneous disruption even in the absence of bombardment by other particles. For such species, the values of r would continue to change even after freezing-in of the equilibrium processes. This possibility should not be beyond the reach of eventual experimental verification, since the energy difference between those stable nuclei which are found in abnormally large amounts, and the parent radioactive nuclei, may be obtained from atomic reactions studied in the laboratory. The appropriate value of ΔA could then be inserted in the plot, or conversely, if the parent nucleus is known, the value of ΔA could be approximately calculated from the plot, a procedure which may be of value in some cases.

(iii) In the absence of other explanations, outstanding exceptions such as the ratio $^7Li/^6Li$ would point to nuclear reactions capable of maintaining this equilibrium at lower values of neutron activity than the bulk of the processes responsible for the maintaining of detailed balancing in nuclear equilibrium. It is of conside able interest to examine how far such reactions can be verified under laborator conditions.

(iv) It will be noted that the form in which equation (7) has been state deliberately avoids a number of difficulties, whose discussion is of great interes but which introduce unnecessary uncertainties in the conclusions. For example the thermodynamic activity of the neutrons under freezing-in conditions can b calculated from the value of the intercept of the curve in figure 1, but this involve further assumptions, which will not be dealt with here. It is interesting to observe nevertheless, that an increase in the thermodynamic activity of the neutrons unde freezing-in conditions would shift the whole curve upwards (by increasing th intercept of the $-\log r$ axis, algebraically), whereas a change in the freezing-i temperature would merely affect the slope of the curve.

(v) In calculating the numerical value of the freezing-in temperature, n allowance has been made for possible changes in the scale of atomic energies wit time. If differences in energy between isotopes were substantially smaller at th time of freezing-in than they are calculated to be from currently observed mas differences, this would lower the estimated freezing-in temperature, but i would not affect the general use of the plot as a test for abnormal nuclear trans formations. [Cf. references to the work of E. A. Milne and colleagues i Johnson, (1945).]

REFERENCES

Aston, 1942. *Mass Spectra and Isotopes* (London : Arnold and Co.).
Chandrasekhar and Henrich, 1942. *Astrophys. J.*, **95,** 288.
Harkins, 1917. *J. Amer. Chem. Soc.*, **39,** 856.
Johnson, 1945. *Time, Knowledge and the Nebulae* (Faber and Faber).
Lattes and Wataghin, 1946. *Phys. Rev.*, **69,** 237.
Pollard, 1940. *Phys. Rev.*, **57** 1186.
Okuda *et al.*, 1941. *Phys. Rev.*, **59,** 104; *ibid.*, **60,** 690.
Seaborg, 1944. *Rev. Mod. Phys.*, **16,** 1.
Wataghin, 1944. *Phys. Rev.*, **66,** 149.
v. Weizacker, 1938. *Phys. Z.*, **38,** 632.

RUTHERFORD AND THE MODERN WORLD

By M. L. OLIPHANT, F.R.S.

The Third Rutherford Memorial Lecture, delivered 7 *October* 1946

§1. RUTHERFORD THE MAN

CAN we attempt to assess the impact of Rutherford and his work upon the science and life of the present time?

Perhaps it is too early to form any balanced judgment of the value of his work and of his influence on others, but it should be possible, even for one who cannot be dispassionate about a man whom he revered as a scientist and loved as a man, to make some timely remarks upon Rutherford's major contributions to the scientific life and outlook of to-day.

Before considering any details it is instructive to measure his greatness in figures concerning one aspect of the release of atomic energy—an achievement of mankind based almost entirely on the work of Rutherford, his colleagues and his students. During 1945, the electricity undertakings of this country generated 37,281 millions of kilowatt-hours of electric power. If the steam plant employed had the high overall efficiency of 30 per cent this would require the burning of about 18 million tons of coal. If the inherent possibilities of atomic energy can be realized in practice, and many of us are convinced that they can, the amount of "atomic fuel" required, if derived from uranium or thorium, would not be much more than 5 tons. In order that the cost of electricity should be the same for both types of fuel, the prime cost of "burning" nuclear fuels could be 3 million times as great as that of burning coal. It is not unreasonable to believe that without Rutherford such great possibilities would not have been realized for a very long time.

Rutherford's faith in nuclear physics as a field of human endeavour was not shared by all. He was often criticized for failing to work, and to train his students, in the useful aspects of physics which had already found practical application. His greatness is apparent in his steady pursuit of the frontiers of physical knowledge, leaving to lesser men the more obvious work of consolidating knowledge, the broad outlines of which had already been explored. He recognized good work in any branch of physics or of any of the other divisions of science, but he had a special love for his own chosen subject, which he called "A Tom Tiddler's Ground", where anything might turn up and where preconceived ideas and theories often toppled to the ground as new experimental facts were discovered.

§2. RUTHERFORD'S DISCIPLES AND THE WAR

Real fortune favours few. This is as true in physical science as in other walks of life. Those who were able to work as pupils and colleagues of Lord Rutherford are to be numbered among those on whom fortune has smiled. Many who worked with him and knew him well have tried to visualize his reaction to the terrible possibilities which the new things in science make probable, to the intrusion of secrecy into pure science and to the growing demand for a special place in society, with proportionally greater income, for scientists of all kinds. Although we accept the fruits of the new spirit and of the new regard of the world for the scientific wizards whom it fears, at least we know that his view was right and that ours is wrong. For him our compromise would have been impossible.

During the last hundred years there has been a continual succession of great men in British physical science, from Faraday and the giants of the Victorian era to Rutherford himself. Today there is no great figure, no one man who by his work, his teaching and his example, can exert the same influence on science as a whole or upon the government of the country. Perhaps the day of the great individual in physics is past and teams of lesser men will progress faster and faster. However, the lessons of the past show that the landmarks of discovery were the product of rare genius and it is difficult to believe that this will not be so in the future.

The development of atomic energy has made use of the accumulated capital represented by the work of Rutherford. Great as is the material achievement, it has been in fact development work—what is called applied physics—and no new discovery has been made in fundamental physics which is of the first or even of th second magnitude. The last six years have been years of stagnation in pur physics, and there is much leeway to make up if there is to be a worthwhile physic in the future. The present generation of physicists has had no contact with th atmosphere of a research laboratory. It knows only the feverish atmosphere o development for an urgent application for war purposes or of a study of phenomen for a material end. Its salaries have often been as high as Rutherford ever earned and its idea of personal progress has been to secure elevation to a position commanding authority and increased pay. Its ability to recover from this diseased state and to recapture the spirit which animated Rutherford will be a measure o the ability of this country to play a great part in the physical science of the future

When the war began there were two great projects where physicists could play a major part. The possibilities of obtaining atomic energy through the fission process had been made clear by the discovery, by Joliot and his co-workers, tha neutrons were emitted. The revolutionary idea of an atomic bomb of unprecedented power was envisaged early in the war by Frisch and Chadwick. However the problem of separating the isotopes of uranium, which seemed essential, wa obviously one demanding long-term development and involving much uncertainty The imminent dangers threatening this country made the alternative problem o radar defence more attractive to many physicists. In both these fields Rutherford's pupils brought his methods of direct attack to bear, and made great strides very rapidly because they were able, at all times, to build on fundamental principles and were unhampered by standards of practice.

Chadwick, to whom Rutherford's leadership in nuclear physics had naturally descended, played a part in the development of atomic energy which will be fully appreciated only when the whole story is told. In his handling of men who worked with him, in his delicate task of relationship with the U.S.A., his legacy from Rutherford was clearly apparent. Perhaps he alone of the British scientists who served in the war preserved throughout that clarity of physical insight and feeling for fundamentals which was characteristic of Rutherford. Most of us found ourselves greatly affected by the scientific compromise which seemed essential to progress and as a result emerged from the war more interested in gadgetry and quick results than in the intellectual side of science.

The migration of Rutherford's students into radar was a surprising feature of the early years of the war, and here they played a prominent part. Men like Cockcroft, Lewis, Dee, Skinner, Ratcliffe, and many others working with Watson-Watt and Rowe, together saved Great Britain from defeat and contributed greatly to her victory. They introduced into the work a totally different atmosphere, for they worked with their fellows as equals and believed in detailed discussion, even with the junior workers, at all stages. The story of what they did and how they achieved it has been told again and again, but one and all they acknowledge that they merely handed on or used the methods which they had learnt from Rutherford and Chadwick in the Cavendish Laboratory.

Blackett, after playing a part in radar, moved on to the Admiralty, where he brought to a high pitch of perfection the new art of operational research—the application to actual military operations of the methods of scientific reasoning.

With the help of Fowler, E. J. Williams and Bullard, he was largely responsible for the defeat of the U-boats at a crucial period in the war.

In these fields and in many others Rutherford's pupils assured the predominance of this country in science applied to war. Tizard, a close personal friend and great admirer of Rutherford, led a mission to America which handed over to its scientists the fruits of our experience and invention, so that when they joined us in the war they also shared the new applications of scientific method and practised these for themselves.

It is disappointing that the full fruits of this influx of Rutherford's spirit into the government and service establishments are not being retained. It was unlikely that many of the best physicists would choose to remain in employment where secrecy and restrictions applied to their work, but such splendid establishments as T.R.E. might have been preserved as national laboratories, possessing a great deal of freedom from day-to-day routine, with great profit to our country.

§3. RUTHERFORD AND THE EMPIRE

Rutherford never forgot that he was a New Zealander or that his early work was done in Canada. He was always sympathetic towards students from the Dominions who came to work with him. His laboratories in Manchester and Cambridge became the Mecca of those overseas students who were lucky enough to go abroad to study, and particularly for those physicists who came to England under the auspices of the Exhibition of 1851. Here they found Mr. Evelyn Shaw, the Secretary of the Commission, ever ready to help them carry on, and in Rutherford they found the kindly sympathy in personal matters, combined with hard drive and an impassioned belief in experimental science, which brought out all that was good in them. Under his guidance, and that of Chadwick, some of us who must have been unpromising material became reasonably competent physicists who, whatever their limitations in fundamental intuition, at least try to keep alive the spirit of his approach. Some have remained in this country, others have returned to the Dominions, and have enriched scientific life there out of all proportion to their numbers. Marsden in New Zealand and Schonland in South Africa are responsible for their countries' science as a whole, while in the Universities and other scientific institutions Rutherford's students hold a surprising proportion of important posts. In the Dominions, as in this country, these men played outstanding parts in their countries' efforts when war came.

Rutherford's Dominion birth and sympathies gave him also a passionate belief in the British Commonwealth of Nations. He seemed always to think of the Commonwealth as a unit, the existence of which, and the predominance of which, he never questioned. This reacted not only on men from overseas but also on his English students and colleagues, so that among them there is a greater appreciation of the problems of the Commonwealth than is common among the citizens of any part of the Empire. Yet he was a true internationalist, ready to give credit for achievement whatever the origin, colour or creed of the worker. At the time of the German persecution of the Jews he was foremost in his defence of their claim on us and worked indefatigably for their cause. He gave to many a home in his laboratory and helped others to secure positions abroad.

Rutherford's faith in the British Commonwealth and in international co-ope ation became the faith of many of his students. It is difficult, for instance, fo these men to understand why in great projects, such as atomic energy, the problem is not tackled co-operatively within the Commonwealth, whose different part could contribute in large measure to the success of a venture which may mean fa more to them in its peaceful aspects than it will ever mean for the United States.

Rutherford admired greatly the experimental genius of Faraday and wa intimately familiar with his diary. In his opinion Faraday was the greatest o experimental physicists. When a full assessment is made of Rutherford's own work and of his influence on physics I think it likely that he will rank with Farada as an experimentalist, while the inspiration he gave to his students and collaborator will place him above Faraday in the sum total of his contributions to science Faraday's work laid the foundations of electrical engineering; Rutherford's is th corner-stone upon which is based the exploitation of atomic energy.

§4. NUCLEAR PHYSICS

Rutherford was not afraid to put forward a hypothesis which was helpful in explaining experimental results because further evidence might prove it wrong He did not hesitate to publish experimental data which, because of the nature of the problem, might turn out to be incomplete or wrongly interpreted. Yet his work and that of collaborators in his laboratory, was singularly free from those hasty and misleading conclusions which have sometimes been published in the insane struggle for priority. His earlier work was extraordinarily painstaking and complete. His discovery, and subsequent investigation, with Chadwick, of artificial transformation by α-particle bombardment has been amplified, but in no case disproved. When, in his laboratory, Chadwick discovered the neutron and Cockcroft and Walton first observed transmutation by artificially accelerated particles, the evidence given was complete and satisfying and started a whole train of fresh work throughout the world. This ability of Rutherford to transmit to others his care and thoroughness, as well as his enthusiasm, has been responsible for the upsurge of great work in nuclear physics in this country and abroad.

Rutherford's strong personal association with Niels Bohr, from the Manchester period onwards, contributed to his proper but balanced appreciation of the place of theoretical physics in the advance of the subject. He refused to be bluffed by the occasional enthusiast who felt that the solution of the appropriate wave-equation could give the answer to all problems of physics. To the end of his life he had an almost fanatical belief in the power of the experimental method. The discovery of such unsuspected phenomena as the fission of uranium and thorium, leading to the atomic bomb, was a striking example of the correctness of his point of view. In fact theoretical advance is due as much to the experimental discovery of new facts and new laws as to advances in theoretical technique, and at present fundamental theory is waiting upon experimental results which would have been available but for the war. Little progress is to be expected in the theoretical physics of the nucleus until data are available on the scattering laws for protons and neutrons at energies in excess of 100 Mev.

Rutherford believed that because of the inevitable lag between academic discovery and practical application it should be possible to foresee, to a usefu extent, the general trend of application of science to industry, and in that way to provide in advance for new things, thus avoiding dislocation of economy and reactionary fighting against progress in technology. In one of his rare speeches in the House of Lords he advocated the setting up by the government of a "Prevision Committee", consisting of scientists and others, who could advise on economic changes which might result from existing scientific knowledge. So far as I know this very practical suggestion has not been implemented, but clearly it will become more necessary as science is applied more vigorously to our economy.

An academic atmosphere, with complete freedom to follow whatever paths of investigation seem desirable or necessary, was, according to Rutherford, an essential for progress in fundamental science. He felt it to be a grave mistake for a man with high abilities in academic physics to be tempted into industrial or State laboratories. I remember an occasion when the director of scientific research of one of the service departments asked Rutherford whether he had one or two good men who could join his division. Rutherford's immediate reply was "Look here, So-and-So, I know the conditions in your place; do you seriously think if I had a good man I would send him to you?" Perhaps such an attitude does not help to improve the standards of the scientific services, but it is certain that the men he refused to send at that time had a far freer hand and made a greater contribution to victory when they joined the Services with fresh minds after war had begun than would have been possible under the conditions applying to Service establishments at an earlier date. When I became an Assistant Director of Research in the Cavendish Laboratory, Rutherford talked to me about the choice of investigations for the research students. "You know, Oliphant, in this game it is rather important to choose the right experiments to do, but it is even more necessary to know when to stop." He added that in industrial and government laboratories men were often assigned to a problem, or class of problem, and had to work in that field till they retired. He believed that no man could make creative contributions to a subject through a particular line of attack for more than a few years. After that he became stale. Yet, despite his reluctance to see the really original mind in physics leave the fundamental for the applied side of the subject, he was no intellectual snob. He recognized and admired achievement in applied physics and gave credit to industry for the tools and devices its work provided for his own experiments.

I have indicated already that the development of atomic energy is likely to rank in the modern world with the birth of electrical engineering in the 19th century. Recent very conservative reports by engineers in America give a very optimistic picture of the possibilities of providing industrial power on an economic basis by "burning" uranium in a modified form of the relatively inefficient type of "pile" operating at Hanford. They conclude that on the same assumptions with regard to amortization of capital charges and allowance for running-cost, electricity generated from atomic energy would cost in the U.S.A. about 30 per cent more than electric energy derived from coal, and they point out that coal costs are rising steadily whereas the cost of atomic energy will undoubtedly decrease as development proceeds. On this basis electrical energy could already be derived from uranium in this country at a cost less than that from coal-burning stations; some of us believe that with reasonable development the cost will be considerably less.

Devons has pointed out that it is probably unimaginative to think of atomic energy merely as a substitute for other forms of energy. The great difference in kind and in concentration of the energy will mean that it will be applied in totally new ways for totally new purposes. The atomic age will arise not so much from a replacement of existing fuels by fissile materials as from new processes which now become possible.

It helps to understand the greatness of Rutherford if we examine the outstanding events, the landmarks, in the unravelling of our understanding of the nucleus, which has culminated in the release of atomic energy. In that way we can see clearly how firmly it all rests on his work; on the intuition of his genius.

The nature of radioactive change

After the discovery of radioactivity by Becquerel and the isolation of radium by the Curies, the further development of our knowledge of radioactive change came about almost entirely from Rutherford's work. The natures of the three types of radiation were recognized by him and named α, β, and γ radiations. Brilliant experiments proved that the α particles were charged atoms of helium. The nature of the ionizing effects of the radiations was developed and the energies deduced. With Soddy he put forward the conceptions of radioactive series and of isotopes.

The nuclear model of the atom

The α particles were always special favourites of Rutherford. He was intensely interested in all their properties and spoke of them as though he knew their colour and even their characters and idiosyncracies. It was natural that he should investigate in detail what happened when these particles passed through matter. For this purpose he developed the use of the scintillation method for detecting single α particles, a method discovered by Crookes, and with Geiger he invented the electrical counter which has played such a large part in modern work. The deviations from straight paths which were occasionally observed led to the careful investigations of scattering and of the scattering laws. On the assumption that the forces between the charged α particles and other atoms were electrical in nature, Rutherford and his colleagues were able to show that the closest distance of approach of the energetic α particles to the centre of electric force in the nuclei of the struck atoms was far smaller than the radius of the atom itself. This led to the modern picture of the atom as a central core, or nucleus, in which is concentrated practically the whole of the mass, carrying a positive charge equal to the atomic number of the atomic species, and surrounded by the number of orbital electrons necessary to make the atom neutral.

Electrons, rotating about the nucleus, were not stable, on the hypothesis that the laws of mechanics governed their motions, as they would radiate energy and fall into the nucleus. Bohr, at that time working with Rutherford in Manchester, found a solution to this problem by the bold conception of stationary quantum states, and the nuclear model of the atom was securely launched.

Artificial transmutation of atoms

During the work on the scattering of α particles it was observed that fast protons were produced. These protons could have arisen from collisions between the

α particles and hydrogen present as an impurity in the bombarded substances. The transfer of momentum which occurs in such collisions would give to the protons a definite maximum energy, actually 16/25 of the energy of the α particles. Examination of the ranges, and therefore of the energies of the ejected protons, showed that in most cases they did arise in this way. However, in the case of nitrogen Rutherford observed that protons were produced with energy considerably greater than this maximum. He concluded that the only possible explanation was to assume that the nitrogen had undergone a nuclear transformation in the collision, that the α particle remained inside the struck nucleus and a proton was ejected, the nitrogen being transformed into oxygen:

$$^{14}_{7}\mathrm{N} + ^{4}_{2}\mathrm{He} \rightarrow ^{17}_{8}\mathrm{O} + ^{1}_{1}\mathrm{H}.$$

In this reaction it was assumed that the proton carried away the excess energy available in the reshuffling of the nuclei.

This was the first observation of the artificial transformation of one element into another. Blackett, in Rutherford's laboratory, confirmed these assumptions by observing the reaction in the expansion chamber, showing that momentum was conserved but that energy was not.

An extensive series of observations was then made by Rutherford, with Chadwick and others, and it was shown that several of the light elements, the nuclei of which could be penetrated by the α particles available, could undergo similar transformations. Experiments were afterwards carried out elsewhere, notably in Vienna, but in general the results were not reliable, and the evidence accumulated in the Cavendish Laboratory remained unchallenged.

In the course of this work it was observed that beryllium when bombarded with α particles gave rise to a very penetrating type of radiation which was thought to be high-energy gamma-radiation. This radiation was investigated in some detail in the Cavendish Laboratory, in Germany and in France, and it was found to be difficult to reconcile its properties with those of other forms of γ radiation. Joliot and his wife, the daughter of the discoverer of radium, found that the radiation from beryllium was able to project protons from hydrogen-bearing substances in a manner similar to the projection of electrons by X rays in the Compton effect, but they found it very difficult to devise a picture of the process which would obey the laws of conservation of momentum and energy without assuming a prohibitively large energy for the γ ray. It was difficult also to account for the large probability of ejection on the assumption that the radiation was electromagnetic in character.

Discovery of the neutron

Chadwick returned to this problem in the Cavendish Laboratory, and in 1932 discovered that the radiation from beryllium was able to project atoms of all kinds, the maximum kinetic energy of which varied in a regular way with the mass of the struck atoms. He concluded that it was impossible to account for these results on the assumption that the radiation was electromagnetic in character, and he put forward the hypothesis that beryllium, when bombarded with α particles, emitted a new kind of particle which carried no electric charge. He was able to calculate the mass of these "neutrons" from the variation of the energy of various atoms which recoiled in collisions, and showed that it was about the same as the mass of the proton

In his Bakerian lecture to the Royal Society in 1920, Rutherford discussed the properties of a neutral particle of mass unity. "Under some conditions it might be possible for an electron to combine . . . . closely with the H nucleus, forming a kind of neutral doublet. Such an atom would have very novel properties. Its external field would be practically zero, except very close to the nucleus, and in consequence it would be able to move freely through matter. Its presence would probably be difficult to detect by the spectroscope, and it may be impossible to contain it in a sealed vessel. On the other hand it should enter readily the structure of atoms, and may either unite with the nucleus or be disintegrated by its intense field, resulting possibly in the escape of a charged H atom or an electron, or both . . . . The existence of such atoms seems almost necessary to explain the building up of the nuclei of heavy elements . . . ."

It is clear that Rutherford understood to a remarkable degree the importance of a neutron as a fundamental particle in the structure of nuclei, and he knew what properties such a particle must possess. It was natural that an intensive search should be made in the Cavendish Laboratory for evidence supporting this hypothesis, but this search was unsuccessful until the advances in counting technique by electronic methods which were due to Wynn-Williams, also working in the Cavendish Laboratory, made it practicable to detect the nuclei recoiling elastically from neutron collisions.

With this background of expectation, Chadwick's discovery of the neutron followed naturally from the years of intense study of nuclear phenomena which Rutherford and he had spent. Chadwick's paper describing the discovery is a model of completeness. There remained no doubt whatever about the existence of the new particle or of its properties.

Transmutation by artificially accelerated particles

Rutherford became interested in the possibilities of producing fast particles with which to bombard nuclei, for he realized that the energy, kind and number of bombarding particles available from radioactive substances were very limited. New information might be obtained by using particles other than α particles, and effects which were too weak to observe might become appreciable if much stronger sources of bombardment were available. Accordingly Allibone, and later Cockcroft and Walton, developed in the Cavendish Laboratory techniques for producing high voltages and for applying them to evacuated accelerating tubes. In particular, Cockcroft and Walton accelerated protons derived from an electric discharge through hydrogen and bombarded targets of the light elements. Although the potentials which they had reached were only 400000–500000 volts, it was felt worth while to try since, according to the newly developed wave-picture of the penetration of particles into nucleus, detectable effects might be anticipated for the lightest elements.

These experiments gave a successful result in 1932. It was found that the penetration of protons into lithium gave rise to a copious emission of α particles. This important observation has had a tremendous effect on physics, for it started the whole modern era of nuclear physics. The reaction may be written:

$$^{7}_{3}\mathrm{Li} + ^{1}_{1}\mathrm{H} + \text{k. e. of proton} \longrightarrow ^{4}_{2}\mathrm{He} + ^{4}_{2}\mathrm{He} + \text{energy},$$

and if we substitute the known masses of the atoms involved, including the mass-equivalents of the energies, we find:

$$7{\cdot}0182+1{\cdot}0081 \rightarrow 4{\cdot}0039+4{\cdot}0039+0{\cdot}0185.$$
$$8{\cdot}0263 \rightarrow 8{\cdot}0263.$$

This remarkable equality proved quantitatively for the first time the exact validity of the Einstein relation $W = M.c^2$. Several other light elements were shown to undergo transmutations when bombarded by protons. Rutherford and others in the laboratory followed up this pioneering work by a careful analysis of the effects produced by bombarding substances, especially deuterium itself, with ions of deuterium, or deuterons. This last series of observations was carried out with less than one-fifth of a cubic centimetre of heavy water provided from his first production by Professor G. N. Lewis of California.

This work was rapidly extended in other parts of the world, especially by E. O. Lawrence in California, using the powerful cyclotron method of acceleration which he had developed. It became the fashion in physics to bombard atoms, and all over the world equipment grew up to enable experiments of this sort to be carried out. However, although many more reactions were studied in detail and far more powerful effects were observed elsewhere, the pioneering work in Rutherford's laboratory remained the supreme example. No discoveries of importance approaching these early observations were made elsewhere.

Artificial radioactivity

The Joliot-Curies, in Paris, had discovered that certain substances could be rendered radioactive by bombardment with α particles. The radioactive materials were unstable isotopes of the normal elements, which transformed by emission of positive or negative electrons into stable isotopes of neighbouring elements. Cockcroft and Walton showed that these radioactive forms of the elements could be prepared by their method of bombardment with protons or deuterons, and this rapidly became the standard method for preparing strong sources, especially where cyclotrons were available. These artificially radioactive substances can now be produced in almost unlimited quantities as a by-product of atomic energy, and will soon be available in this country, as in America, for use in chemistry, biology and medicine as "indicators" of the movements and history of any element in any compound.

Transformations produced by neutrons

Feather, in the Cavendish Laboratory, first showed that neutrons could produce nuclear transformations in exactly the same way as protons or α particles, but the major early work on transformations produced by neutrons was done by Fermi and his collaborators in Rome. In particular it was shown that neutrons could be slowed down to thermal energies by multiple collisions with hydrogen in such substances as water or paraffin wax, and that these slow neutrons were particularly effective. They were able to enter most nuclei and were often captured there, producing unstable radioactive isotopes of the same element. In some cases the probability of capture was extremely high due to the existence in certain nuclei of energy levels enabling resonance transitions into the system by neutrons with small kinetic energy.

Fermi examined the reactions produced in most elements. The neutron being uncharged, is able to enter heavy nuclei of high atomic number as easily as those of small mass and charge, so that in contrast with accelerated protons and deuterons strong effects were observed with the heaviest elements. His observations with uranium led him to believe that by electron emission after neutron capture, elements of greater nuclear charge than uranium, the so-called transuranic elements, could be formed, but some of the evidence was not very satisfactory.

These observations were followed up in Berlin by Hahn, who had been one of Rutherford's earliest research students in Montreal. Using chemical methods of separating the radioactive products of the reaction of slow neutrons with uranium, he was able to show, in 1938, that these were often elements of medium atomic weight, like barium and strontium. He therefore put forward the hypothesis that uranium, after absorption of a neutron, could undergo a new type of transformation by splitting into two approximately equal parts—a "fission" process, which was accompanied by an enormous release of energy. This hypothesis was immediately confirmed in Copenhagen and in America. Shortly afterwards it was shown by Joliot and his collaborators that in the fission process several neutrons were emitted. At once the possibility of producing a chain process became apparent, and the search began all over the world for methods of achieving such a divergent process, in order to produce both industrial power and atomic bombs.

Surely, after consideration of the process by which our knowledge of nuclear physics has advanced to the stage where atomic energy is available, we must admit that of all men Rutherford must stand out as the pioneer of the new age. His was the genius which gave it birth.

The future of nuclear physics

Nuclear physics, as a subject for academic research, faces a serious crisis in all countries. The grave military implications of atomic energy are bound to mean restriction on the field of activities in nuclear physics which can be permitted without control and supervision. It is fortunate that the sections of the subject which need special supervision are less those dealing with the frontiers of our knowledge than those concerning the detailed accumulation of information about particular types of nuclear reactions. This means that academic nuclear physics can still be carried on with some considerable freedom in the more fascinating fields of the subject provided only that the law relating to control of atomic energy is wisely interpreted and administered.

§ 5. CONCLUSIONS

Rutherford's greatest contributions to our knowledge of the structure of matter came from his intensive investigations of the scattering of α particles—his use of these atomic projectiles as probes with which to examine the fields of force within the atom and even within the nucleus itself. The power and elegance of this method is obvious, and it is clear, too, that much fresh knowledge is to be expected from the use of particles such as protons and neutrons, in place of α particles, especially if much higher energies can be employed. The meson theory of binding for the constituent particles of the nucleus, while providing the best picture we have of the nuclear field, is far from satisfactory, and much more information is required about free mesons, as well as about the laws of force between elementary particles at very close distances of approach. To create a pair of mesons it appears that elementary particles with energies in excess of about 300 Mev. are required, while the scattering of particles of these energies could give information about the laws of force. It seems likely, therefore, that efforts will be made to produce protons and electrons with energies of 1000 Mev. or more. Several methods for achieving this have been suggested, and if the technological difficulties can be overcome we can face with confidence a new era of increasing knowledge of that fertile "Tom Tiddler's Ground" where Rutherford dug so well.

DISCUSSION

on paper by E. W. H. SELWYN and J. L. TEARLE entitled "The performance of aircraft camera lenses" (*Proc. Phys. Soc.*, **58**, 493 (1946)).

Mr. J. W. PERRY. Mr. Selwyn and Dr. Tearle have contributed considerably to photographic optics by this paper. The method employed, considered in relation to the title of the paper, calls for some comment, however, as it is possible to misunderstand its significance if not viewed in the proper perspective. What is presented is not a solution of the problem of photographic lens testing in any fundamental sense, but an empirical bridge to short-circuit important branches of the subject which would have been obstacles to an early fruition of the work. The investigation is, in fact, a typical example of an empirical treatment of an involved subject rendered justifiable by an emergency. Thus one should not be misled by an undoubted success of the work into regarding the results as ultimately valid, for investigation would certainly be needed to put the practical generalizations thus won, on to a fundamental basis and to free them from restrictive conditions and enforced assumptions. For the data upon which the work is based refer only to a certain definite epoch in the course of the development of the photographic objective. They assume certain commercial types of a mixed kind conditioned by different economic and other arguments affecting their form and complexity. They exclude the aid to correction which non-spherical surfaces can contribute. Even the glass situation will have affected one or more of the objectives assumed. Thus there was no fortunate conspiracy of circumstances in the provision of the data upon which Mr. Selwyn and Dr. Tearle had to work.

It is clear from the results that, apart from any possible defects in the photographic plate, the performance of the lens and plate combination suffers from imperfections in the lens, and that with successive improvements in lens design and in methods of manufacture and of test, steady removal of these and approach to the ideal should result in a general improvement in photographic performance, so far as the lens is concerned. But what is the ideal ? That cannot, of course, be learned from an empirical investigation in which varying physical conditions, instead of being isolated for individual study are confused by other unrelated variables coexisting and simultaneously operative. For this the lens must be considered from the point of view of its immediate function and physical effect upon waves of light. The lens testing interferometer provides precisely this information for lenses of any kind, aircraft camera lenses of course included.

The investigation goes considerably beyond this, of course, in order to produce immediately utilizable information and gives results of performance of aircraft *cameras* as a whole. By so doing it adds considerably to our knowledge. If it also thereby loses claim to scientific validity in a fundamental sense, that is by no means a valid criticism, for whereas the combination of physics and empiricism which enters into such an investig tion can produce valid results to enable a war to be prosecuted, it obviously cannot ta the place of physical studies aiming at an ultimate analysis of the whole process. Th results of such a valuable investigation will be all the more valuable when they are actuall supported by physical information on the lens performance and on the performance of th photographic plate from the point of view of optical processes.

Mr. G. S. Speak. With regard to the improvements in lenses mentioned by th author it may be of interest to note three outstanding increases in performance which hav occurred within the last two years. A 25″ Ross lens is some 50–60% better than woul be indicated by the formula, a 50″ Ross telephoto is 40% better, and a redesign of a 36 telephoto by Wray (original design by Booth) is also 40% better.

Our practice at the Royal Aircraft Establishment when testing lenses photographicall is not to give the exposure at each point in the field which leads to maximum resolution but to expose on axis for this condition and give the same exposure time for points of axis, since this is what happens when the lenses are used in the normal way. For simila reasons the average resolution is calculated over a rectangular picture area instead of circular area.

Work we have carried out with various test-objects indicate that the probable erro of results obtained with a Cobb object is of the order of 7% and that there is no advantag in decreasing the size difference from one group to the next to less than about 5%. Howlett* in Canada obtained similar figures for the accuracy of resolution measurements.

The type of test-object to be used has been discussed many times before. Muc obviously depends on the information required from the test, but if the test-object is to bear some resemblance to the details which the lens will be required to photograph eventually then for testing lenses to be used in air photography an annular test-object proposed by Howlett has numerous advantages, and may be of use in other work.

Finally, I should like to state my opinion that from the point of view of the optica designer resolution tests alone are probably of little use. They should be supported by measurements of aberrations, so that the designer may obtain an idea of the corrections which are most conducive to a good performance. For this reason our reports on lenses for some three years at the R.A.E. have contained full details of resolution tests and measurements of most of the aberrations in the system under test with the idea of placing optica design on a more quantitative basis than it has been in the past.

Mr. G. C. Brock. This work on photographic resolving-power has been described by a previous speaker as "ephemeral". I think that is probably a sound judgment on a long view and having regard to fundamental aspects, but it does not alter the fact that we have witnessed a very notable advance in photographic optics which would not have occurred without the knowledge accumulated in the course of these tests. In 1940 the Air Photography Research Committee were greatly concerned to improve the resolving-power of aircraft cameras, and realizing that more fundamental investigations would take far too long, sponsored this programme of resolution-tests in the Kodak Research Laboratories. The essential first step in any scientific investigation is to establish some basic facts, and at the time no facts were available with which to answer two questions :—

(*a*) Were lenses or emulsions most in need of improvement ?

(*b*) What was the relation between angular resolving-power and focal length ?

I feel quite strongly that we should not forget how limited was our knowledge of these things in 1940. The results of these investigations by Dr. Tearle and others substituted definite facts for a mass of conflicting opinions and paved the way for the advances made by our opticians in the construction of narrow angle reconnaissance lenses. We should now like to see a similar improvement in wide-angle lenses.

This work has of course made a great contribution to the advances in the general theory of photographic resolving-power, one of the most interesting aspects being the great difference between the photographic and visual resolving-powers of a given lens, and the apparent lack of correlation between them.

---

* Howlett, L. E., *Journal of N.R.C. of Canada* (July 1946).

I should also like to emphasize that in work on air photography, where resolution of ground detail is affected by other things, such as image movement, a knowledge of the resolving-power of the lens/film combination has been an indispensable condition of progress.

The resolution test can be criticized from many angles, and is funadmentally a temporary device which we must hope to see replaced or at any rate extended by some more absolute method involving a measurement of the intensity distribution in the image plane. Nevertheless it has done excellent service over the past few years and we will be returning to it quite often for some time to come.

Authors' reply. We have shown in the paper that the effect on the photographic resolution of the decreasing illumination from centre to edge of the field is very small in the systems investigated, and the method of exposure is therefore of little importance. This result was, of course, unknown at the outset of the work, but was available to the Royal Aircraft Establishment when photographic lens testing was commenced there on a large scale. In lenses where the decrease in illumination with increasing separation from the axis is marked, some modification in the method employed at the Royal Aircraft Establishment would be necessary ; it would, for example, be preferable to give the optimum exposure at some point between the centre and the edge of the field, allowing the axial region to be over-exposed and the peripheral region to be under-exposed.

REVIEWS OF BOOKS

Photoelectric Cells, by A. Sommer. Pp. 104. (London: Methuen and Co. Ltd., 1946.) 5*s*.

This latest addition to Methuen's Monographs on Physical Subjects gives exactly the kind of information needed for choosing intelligently what kind of emission cell to use as a tool in a particular research, or as a component in apparatus designed for a specific purpose. It also fills the gap between the more advanced text book on photoelectricity, and the standard works on photocell applications which give rather scanty attention to the cell itself. It may therefore serve both as a *vade mecum* for the experimenter in other fields and as a simple introduction to the subject for the non-specialist. It should be noted that it does not deal with barrier layer, or rectifier, photocells.

Many will wish it had been a little larger, and may find in its brevity the reason for a slight distortion of perspective. While the choice of the silver–oxygen–caesium, antimony–caesium and bismuth–caesium cathodes as illustrative examples is good, because they are widely used as well as typical, other cathodes perhaps deserve more than a brief mention, particularly the silver–oxygen–potassium and silver–oxygen–rubidium. For the experimenter, more complete information on the spectral response and thermionic emission of these cathodes, and of others, including those used for the ultra-violet, as well as on the spectral absorption of cell envelopes in ordinary and special glasses, would have been valuable. Also, much greater emphasis might have been laid on the importance of electrode design in cells to be used for purposes of measurement, so as to dispel any impression that one type of cathode is appreciably better than another in this respect. The differences in the metrological performance of different cells now on the market almost certainly result from differences of design, and not of cathode material. The production of a well designed cell is more difficult with some cathodes than with others, but the difficulties are not insuperable. To most scientific workers the photocell is essentially a measuring instrument, and one hopes therefore that manufacturers will not allow these difficulties to impede the regular production, in a form suited for precise work, of cells with a greater variety of cathodes including especially the newer very sensitive alloy cathodes. Furthermore, it is only with cells of such a kind that research on the photoelectric effect itself can give completely reliable results.

The monograph appears to be free from minor blemishes of consequence, though it is not quite true to state (p. 28) that no studies have been made on the spectral absorption of the Ag–O–Cs cathode film. Asao, in Japan, published the results of an interesting stud of this subject in 1940.

The simplified theoretical introduction is good and quite adequate as a backgroun to later chapters devoted to more detailed aspects. The subdivision under headings an sub-headings is clear and useful, and the style is simple and fluent.

Dr. Sommer has succeeded well in presenting his subject clearly and simply to th lay reader possessed of a certain background of general physical knowledge, while th amount of technical information compressed without loss of continuity into so small volume, will undoubtedly commend the book also to a wider circle of users of photocells.

J. S. PRESTON

A Handbook of Telecommunication (*Telephony and Telegraphy over wires*), b B. S. Cohen. Pp. xiv+437, with 281 diagrams and numerous tables (London: Pitman and Sons, 1946.) 30*s*. net.

The arts of telephony and telegraphy, particularly the former, have made great stride in the past few decades. An account of the stages of this progress is of interest to th physicist as well as to the telecommunication engineer. This book provides a welcom medium for surveying the new territory. The author, B. S. Cohen, was well fitted to undertake the task of writing it. For many years he was in charge of the research wor in the extensive laboratories of the British Post Office at Dollis Hill. He read severa papers before the Institution of Electrical Engineers, and his name constantly appear in the discussions recorded in their Journal. He gave one of the Faraday lectures, hi subject being " The Long-distance Telephone Call ". Before his premature death in 1940 he had fortunately completed the manuscript of this book. The final revision was undertaken by Mr. F. G. C. Baldwin with the assistance of some of Cohen's colleagues.

That the more striking advances lie in the branch of telephony is reflected in the preponderatingly large space the author has elected to give to it. After an introductory chapter eleven chapters are devoted to telephony, one only to telegraphy, one to the thermionic valve, and one to measurements, the last two being more or less common to the two branches.

Speech and music can now be transmitted over wires almost without limit of distance over the earth's surface, with high fidelity of reproduction at the receiving end. In 1907 the highest indispensable frequency of reception was considered to lie between 800 and 1100 c./s. To-day it is internationally agreed that speech reception shall cover the band 300 to 3400 c./s, and broadcast music the band 50 to 6400 c./s. The improvement in quality is distributed over all stages, in the line and at each terminus. In the year quoted, Bell's receiver was much as it had left the inventor's hand in 1876. Today its frequency response curves are much nearer to the ideal, and the sensitivity is appreciably greater. In 1907, Pupin's loading coil was just coming into general service. In the quarter century that followed, several million pounds sterling had been invested in these coils in this country. By 1936 they were fast disappearing from the lines, to reappear in the terminal apparatus.

The introduction of the thermionic valve solved the problem of distant transmission. Black's 3-valve negative feed-back amplifier met the requirements in a highly effective manner. By its insertion at suitable intervals along the line the attenuated power is stepped up by 50 or more decibels at each stage, whilst close fidelity is preserved over the whole range of frequencies.

The great expansion of business over trunk lines raised acutely the problem of economical transmission. Here the adoption of the technique of radio is the remarkable feature, and the co-axial form of air-cored cable proves peculiarly appropriate to the purpose. Using, for instance, carrier-frequencies with an over-all range of 0·5 to 2·1 Mc./s., and allowing a band of 4000 c./s. for each message (only one side-band is transmitted, the carrier and the second side-band being suppressed) it is possible to send 400 independent messages simultaneously over a single pair of wires. The re-introduction of the carrier-frequency at the receiving end involves synchronization of the oscillator at that end to one part in two millions ! Band-pass filters play an essential part in finally resolving the messages.

In telegraphy, too, the carrier principle is employed to the same end, to provide the voice-frequency multi-channel telegraph system. Here the carrier frequencies lie within the audio range. The teleprinter is now in general use, and the vogue of the Morse code is rapidly becoming obsolescent.

These and other notable advances will be found described in some detail in Cohen's book. The task of selection of subject-matter amid the avalanche of change must have proved a delicate one. The reader has, however, been provided with copious references to text-books and original papers, a list of these being appended to each chapter. (The omission of Dr. Mallett's *Telegraphy and Telephony* from these lists is surprising).

A good index is especially valuable in a book of this kind. The index supplied, though running to 15 pp., is perhaps, hardly adequate. Our use of it has disclosed several omissions, and we would suggest that the index be considerably extended in a second edition.

The book, with its descriptions of the latest forms of microphones and telephone receivers and other apparatus, and its clear accounts of the application of physical principles in such important practical fields, will we think be warmly welcomed by teachers and students of physics, and should find a place in all scientific libraries. D. O.

Understanding Microwaves, by VICTOR J. YOUNG. Pp. 385 + xi. (New York: John F. Rider Publisher, Inc., 1946.) $6.00.

The first point to understand about any scientific or technical book is its title, and in this connection the reviewer has always been a little doubtful as to the meaning of the term " microwave " and the portion of the radio frequency spectrum to which it applies. In an earlier book entitled *Microwave Transmission*, by J. C. Slater, the first sentence of the preface states that " Microwaves are electromagnetic waves of wave-lengths that we may take, for definiteness, to be between 1 centimeter and 1 meter." In the absence of any specific information to the contrary we may accept this definition as applicable to the work under review, although it is perhaps doubtful if the author intended to be limited at the short end to a wave-length of 1 centimetre.

While it is not clear to what class of reader this book is addressed, it is to be noted that much of the latter portion of the book deals with microwave terminology and contains some relatively advanced mathematical conceptions which are in marked contrast to the elementary nature of the introductory chapters. In such a rapidly advancing subject as radio, it would seem unnecessary for modern books to recapitulate the elementary principles of electricity and magnetism, alternating current circuits and transmission lines, and yet the first five chapters of this book are confined to an exposition of these matters. Chapter 6 deals with Poynting's Vector and Maxwell's Equations; and here, as throughout the book, the mathematics is purposely reduced to a minimum, and the information is conveyed in a graphical and descriptive manner. On page 123 an incorrect formula is given for the complex dielectric constant of a conductor, which has the effect of making this constant directly proportional to the incident electric field instead of being independent thereof, and directly instead of inversely proportional to the frequency.

Some wave-guides and their properties are treated non-analytically in Chapter 7, and resonant cavities are described in Chapter 8, together with some applications. These two chapters give a reasonably clear picture of the basic principles involved, though the difficulty of presenting a concise and really adequate account without the use of mathematics is perhaps more apparent here than in other parts of the book. The following chapter contains a brief description of various types of centimetre wave antennae, including parabolic reflectors, horns and corner reflectors. In a short note on propagation at the end of this chapter reference is made to interference between direct rays and those reflected from the earth, in which it appears that the word interference is used to cover only those cases where the signal received is less than the free space signal. The term " interference ", however, covers the whole phenomenon and embraces both the " bright and dark " bands of the radiation field pattern. Chapter 10, entitled " Microwave Oscillators ", is probably the best chapter in the book. Although admittedly in keeping with the descriptive and rudimentary style of the rest, it nevertheless, within the limitations imposed by such a treatment, gives a clear picture of the manner of operation of klystrons and magnetrons, the valves which play such an important part in centimetre wave technique. The final chapter in Section I of the book gives an outline of the factors which limit the range and application of microwaves in radar and communications. Receiver sensitivity and bandwidth, antenna gain and attenuation during propagation are discussed in relation to this problem.

Section II, comprising the last 100 pages of the book, is devoted to a glossary of microwave terminology and represents a quite useful collection of definitions and theorems applicable in radio technique. It might be noted that (on page 297) the functions referred to should be Hankel and not Henkle. Rather than use the definitions given on pages 309 and 310 it is better to class Fraunhofer diffraction as that in which the radiation source and location of the pattern are each effectively at infinite distances from the diffracting obstacle whilst Fresnel diffraction refers to the case of finite distances from the obstacle. " Line-of Sight Range " is here defined as the maximum distance over which microwaves can be transmitted. Neglecting the case of anomalous super-refraction, this is hardly a good term to use under normal refraction conditions; for radio waves are refracted to a greater extent than light waves in the normal atmosphere, so that, in fact, although a radio receiver might be within the direct range of the transmitter, they need not be optically intervisible.

The author is well aware of the difficulties of the units question and uses in the main the Gaussian system, although occasionally other systems, notably the M.K.S., are applied. It might have been preferable to have used a uniform system throughout.

In general, the book appears to contain a large amount of unnecessary material, while at the same time insufficient space is devoted to matters peculiar to very short wave technique; the portions dealing with antennae and with the propagation of waves could have been expanded with advantage.

R. L. S.-R.

A First Course in Mathematical Statistics, by C. E. WEATHERBURN. Pp. xv + 271. (Cambridge: The University Press, 1946.) 15*s*.

This book, one of the very few which treat of statistical theory in the form which it has attained in the last 30 years, is very definitely a companion to Fisher's *Statistical Methods for Research Workers*. The latter gives, without many concessions to the weaker brethren, a critical statement of the methods to be applied in testing significance and in analysing experimental results. Professor Weatherburn's book, on the other hand, gives the basic theory leading to the methods described by Fisher—theory which for the most part has had to be read in the original papers of " Student ", Fisher and other workers. It is true that a few good text-books have appeared in recent years in which the theory is accurately set out, but none of them could be described as suitable for the beginner.

Professor Weatherburn has therefore rendered a distinct service by writing this book, in which the reader is led as far as an introduction to the analysis of variance, and to the topic of multiple correlation. Before this, the reader has been introduced to small-sample theory and to the χ^2 test of goodness of fit. Much of this involves proofs of the distributions to which various types of statistic tend, and hence has been preceded by descriptions of different standard distributions, the whole introduced by two chapters in which the general notion of a frequency distribution and its chief parameters has been introduced, and the relation of a probability distribution to a frequency distribution explained. From a fundamental point of view this is, perhaps, the most important step in the whole subject. It is here treated well, though very concisely.

From this summary it might appear that, with the exception of chapters 1 and 2, the book merely duplicates Fisher's work, but there is in fact a great difference, and the two are, as the author claims, complementary. Weatherburn gives the theory (in elementary form) and illustrates it with examples in which pains have been taken to keep the actual arithmetical work within bounds. The book does indeed form a good introduction to the theory, and introduces methods, like the moment-generating function and the cumulative functions, which Fisher, in his text-book, has no need to mention. In addition to exercises for practice, each chapter has a set of references for further reading, which the student is advised to peruse while he is working through the book. These should be of considerable value—there is a great deal to be gained by reading the same thing in different words or in a different order—though at first sight the lists seem rather intimidating, owing to their length.

The book is beautifully produced, and seems remarkably free from misprints.

J. H. A.